DEVELOPING A NOVEL HYPOMETHYLATED POPULATION AND CHARACTERIZING A STABLE EARLY FLOWERING EPIMUTANT IN STRAWBERRY (*FRAGARIA VESCA*)

A Thesis Submitted to the College of Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

In the Department of Plant Sciences

University of Saskatchewan

Saskatoon

By

Jihua Xu

PERMISSION TO USE

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Plant Sciences
College of Agriculture and Bioresources
51 Campus Drive
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5A8, Canada

ABSTRACT

DNA methylation, as one of the epigenetic marks, can be passed from generation to generation. Epigenetic regulation of phenotypic traits can be demonstrated by altering methylation patterns with DNA demethylating reagent 5-azacytidine (5-azaC). Woodland strawberry (*Fragaria vesca*, 2n=2x=14) has a rapid generation time (3.5 months in the greenhouse), small plant stature and small sequenced genome size (~240 MB). *F. vesca* makes an ideal model system for genomics analysis and gene functional study in the Rosaceae family.

In order to obtain a *F. vesca* population with the same genetic background but increased variation for epigenetic information, 5-azaC was used to generate a hypomethylated population by treating highly inbred seeds from Hawaii 4 (H4S8). The uniform genetic background was first confirmed using Amplified Fragment Length Polymorphism (AFLP). Additionally, whole genome sequencing demonstrated the epimutagenic, not mutagenic, effect of 5-azaC resulting in primary DNA sequence changes.

Subsequent detailed phenotypic assessments measured flowering time, rosette diameter and stolon emergence time. Expanded variation in these phenotypic traits was observed in the hypomethylated population. Distinct DNA methylation patterns were detected using Methylation Sensitive Amplification Polymorphism (MSAP) in early flowering, late flowering, and small rosette diameter lines. The inheritance of methylation profiles were confirmed by bisulfite sequencing of targeted regions. The inheritance study following early flowering, late flowering, and late stolon emergence time variants through meiosis were demonstrated. It revealed early flowering was stably inherited for at least five meiotic generations but the late flowering phenotype reverted back to the control after two generations. Furthermore, the clonal transmission of the early flowering trait to daughter plants was also confirmed through mitosis.

Characterization of the fifth generation (H4S13) early flowering lines was performed using high-resolution methylation combined with RNA-Seq. Differentially expressed genes involved in the flowering time pathway were detected, especially those involved in the photoperiod pathway. The up-regulation of downstream genes in flowering pathway *NUCLEAR FACTOR Y, SUBUNIT B2* (*NFYB2*), *FLOWERING LOCUS T* (*FT*, the highest fold change), *FRUITFULL* (*FUL*) *and SEPALLATA3* (*SEP3*) likely contributed to the floral transition. Significantly different CG methylation levels of the FT gene and coding sequence

regions were observed. The overall DNA methylation rate at CG (63%), CHG (36%), and CHH (1%) sites based on single base resolution is first reported in strawberry. In addition, a novel MSAP method was established using the isoschizomers *Tfi* I / *Pfe* I to detect DNA methylation at CG, CHG, CHH sites based on restriction sites. Further application of this method into the early flowering lines found the maintenance of symmetric CG, CHG methylation over meiotic generations.

This project for the first time developed a hypomethylated *F. vesca* population, comprehensively studied DNA methylation variation of quantitative traits, and the inheritance of these traits through meiosis and mitosis. These data add to the growing evidence that altering epigenetic variation can be a mechanism for generating increased phenotypic diversity and can be subject to selection. This research indicates the potential of crop improvement through epigenetic modification without changing DNA sequence, avoiding the issue of Genetically Modified Organisms.

ACKNOWLEDGEMENTS

First of all, I wish to express my deepest gratitude to my supervisors Dr. Karen Tanino from University of Saskatchewan and Dr. Steve Robinson from Agriculture and Agri-Food Canada. They opened the door to such an interesting project and guided me with patience, encouragement, and understanding to this field. Their rigorous scientific attitude and extensive knowledge were essential for the success of this research. I would like to express my appreciation to the thesis committee Dr. Yuguang Bai, Dr. Bob Bors, Dr. Gorden Gray, Dr. Hong Wang, and external examiner Dr. Carl Gunnar Fossdal for their kind assistance, valuable suggestions and insightful comments in all committee meetings throughout the project.

In this project, lab technicians Tricia Bender and Kyla Horner from Agriculture and Agri-Food Canada provided technical support of the skills used in the lab, thank you for the patience and professional training. Thanks are extended to lab technicians Rensong Liu and Gowribai Valsala, and greenhouse technician Eldon Siemens from University of Saskatchewan for their assistance. Thanks are also extended to my colleagues Jun Liu, Ian Willick, Pankaj Banik, Kaila Sadowski, Elena Benic, and Dr. Prakash Venglat from University of Saskatchewan for their advice to this project.

My thanks also go to Dr. Janet Slovin from U.S. Department of Agriculture - Agricultural Research Service for kindly provided the seeds material. Thanks are extended to Dr. Mouhu from University of Helsinki and Dr. Lester Young from University of Saskatchewan for their kind advice on the gene expression experiment. Thanks are extended to Erin Higgins from Dr. Isobel Parkin's lab, Agriculture and Agri-Food Canada, for the help of all Illumina sequencing runs in this project.

Thanks are extended to my friends James Dawson, Wen Zang, Jia Sun, Bo Gui, Fangqin Zeng, Runfeng Wang, and Dr. Aron Cory for their support and encouragement.

I would like to thank China Scholarship Council, University of Saskatchewan, Agriculture and Agri-Food Canada, and Agriculture Development Fund from Saskatchewan for the support of this project.

Special thanks go to my husband Zhifa Wang, my parents, my sister, and brothers for their huge support and understanding during my Ph.D study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

5-azaC	5-azacytidine
AFLP	Amplified Fragment Length Polymorphism
AMOVA	Analysis of Molecular Variance
AP1	*APETALA1*
AP2	*APETALA2*
CAL	*CAULIFLOWER*
CCA1	*CIRCADIAN CLOCK ASSOCIATED 1*
CDFs	*CYCLING DOF FACTORs*
CIR1	*CIRCADIAN 1*
CO	*CONSTANS*
COL5	*CONSTANS LIKE 5*
CTAB	Cetyltrimethyl Ammonium Bromide
DEG	Differentially Expressed Genes
DMR	Differentially Methylated Regions
DN	Day Neutral
EF	Early Flowering
EpiRILs	Epigenetic Recombinant Inbred Lines
EST	Expressed Sequence Tags
EUE	Energy Use Efficiency
FCA	*FLOWERING LOCUS CA*
FLC	*FLOWERING LOCUS C*
FLD	*FLOWERING LOCUS D*
FPA	*FLOWERING LOCUS PA*
FRI	*FRIGIDA*
FT	*FLOWERING LOCUS T*
FUL	*FRUITFULL*
FWA	*FLOWERING WAGENIGEN*
FY	*FLOWERING LOCUS Y*
GA	Gibberellin
GATK	Genome Analysis Toolkit
GI	*GIGANTEA*
HAP3B	*HEME ACTIVATED PROTEIN 3B*

HPLC	High Performance Liquid Chromatography
LD	Long Day
LD	*LUMINIDEPENDENS*
LF	Late FLowering
LFY	*LEAFY*
LHY	*LATE ELONGATED HYPOCOTYL*
LMI1	*LATE MERISTEM IDENTITY 1*
LS	Late Stolon
MBD	Methyl-CpG-binding domain
MeDIP	Methylated DNA immunoprecipitation
MET1	DNA METHYLTRANSFERASE 1
MSAP	Methylation Sensitive Amplified Polymorphisms
NFYB2	*NUCLEAR FACTOR Y, SUBUNIT B2*
PCoA	Principal Coordinates Analysis
PRR7	*PSEUDO-RESPONSE REGULATOR7*
RAPD	Randomly Amplified Polymorphic DNA
RdDM	RNA-directed DNA methylation
SAM	Shoot Apical Meristem
SEP3	*SEPALLATA3*
SFL	*SEASONAL FLOWERING LOCUS*
SiRNAs	small interfering RNAs
SNP	Single Nucleotide Polymorphism
SD	Short Day
SOC1	*SUPPRESSER OF OVEREXPRESSION OF CO1*
TE	Transposable Element
TFL1	*TERMINAL FLOWER 1*
TOC1	*TIMING OF CAB EXPRESSION1*
VRN	*VERNALIZATION*
WGBS	Whole Genome Bisulfite Sequencing

1.0 INTRODUCTION

In quantitative genetics analysis phenotypic variation is composed of genotype and environmental variation. Over the last two decades it is becoming interestingly accepted that epigenetics contributes to phenotypic diversity, which provides new insight to re-examine the underlying nature of environmental variance raising questions of evolutionary role of these variation (Hirsch et al., 2012; Diez et al., 2014). Epigenetics add another layer of regulation in the plant genomes that can affect gene expression and in turn alters plant growth and reproduction (Feng and Jacobsen, 2011). Genetics, epigenetics, environmental factors, and their complex interactions together determine the phenotypes (Dwivedi et al., 2011).

Epigenetic information is encoded in DNA methylation, post-transcriptional histone modifications, chromatin structure and non-coding RNA molecules, which together can affect allelic expression and thus quantitative traits (Feng et al., 2010). There remain questions regarding the extent and heritability of phenotypic changes induced by the activation or silencing of gene expression caused by epigenetic regulation. Reports suggest that phenotypes including disease states in humans and abnormalities in plant development can be altered in this way (Finnegan et al., 1996; Nakao, 2001). DNA methylation is a well-studied epigenetic mechanism that controls gene expression. The regulatory effects of DNA methylation on gene expression can be direct, acting on gene promoter regions, or non-promoter gene bodies and intergenic regions (He et al., 2011). This can further interact with DNA-binding transcription factors to influence the structure of chromatin changes (Bird, 2002; Hashimshony et al., 2003). Changes in DNA methylation might also act indirectly leading to genetic mutations through transposons activation (Wolffe and Matzke, 1999).

Fragaria vesca (2n=14), a member of the Roseceae family, has taken on greater scientific importance over recent years. Cytological analysis classifies strawberry into distinct groups based on genome size. There are diploid, tetraploid, hexaploid, decaploids, and octoploid strawberry (Bauer, 1992; Folta and Davis, 2006). The seed to seed life cycle of diploid strawberry is 3.5 months and strawberry has evolved the ability to reproduce through sexual and vegetative ways producing achenes (seeds) and runners/stolons, respectively. Due to its simple genome structure, short generation time and propagation patterns, the woodland strawberry (*F. vesca*) has been recognized as the reference plant in the Rosaceae family for genomic and genetic studies (Folta and Davis, 2006; Shulaev et al., 2008). The scientific value of *F. vesca* has increased with the completion of the genome sequence of Hawaii 4

(Shulaev et al., 2011).

The importance of the epigenetic regulation established by DNA methylation can be demonstrated by altering methylation pattern with pharmacological agents such as 5-azacytidine (5-azaC). Such compounds act as inhibitors of DNA methyltransferases, and it prevents DNA methylation at cytosine targets (Christman, 2002). The 5-azaC treatment leads to a randomly distributed net reduction in cytosine methylation throughout the genome that has the potential to alter the expression of numerous genes. This phenomenon has been exploited to increase phenotypic variations in many plants (Sano et al., 1990; Burn et al., 1993; Fieldes, 1994b; Marfil et al., 2009).

In *Arabidopsis*, a population of epigenetic Recombinant Inbred Lines (epiRILs) was generated by crossing a wild type parent and a ddm1 mutant with reduced DNA methylation (Johannes et al., 2009). The epiRILs were used to study epigenetic variation and transgenerational inheritance. The lack of epigenetic resources and the above characteristics of *F. vesca* make it possible and a reasonable strategy to establish an epimutagenized population in strawberry using 5-azaC treatment to study epigenetic variations and the inheritance of phenotypic traits through successive generations. It could be an effective tool to detect potential epigenetic regulation mechanisms and identify the existence of heritable epialleles. For the first time, our studies demonstrated the potential of epigenetic variations regulating phenotypic changes and the heritable selection of induced phenotypic traits in strawberry. This model system will provide new insight in the area of crop improvement through epigenetic regulation, especially for crops showing limited genetic diversity to be modified using traditional breeding methods. The specific objectives of these studies were to:

1. Develop a novel epigenetic resource in strawberry that has an expanded level of phenotypic variation. 1.1. Assess the genetic background in 5-azaC treated population (epimutagenized population). 1.2. Assess the potential for 5-azaC treatment to induce phenotypic variation and to cause changes in DNA methylation in strawberry plants.

2. Study the epigenetic inheritance of families possessing phenotypic traits of interest. 2.1. Assess the inheritance of observed phenotypic variation among the strawberry lines through meiosis and mitosis. 2.2. Detect the DNA methylation patterns of inherited phenotypic lines.

3. Characterize selected epimutant lines showing stable inheritance of a specific phenotype. 3.1. Detect gene expression levels of epimutant lines possessing a certain phenotype. 3.2. Detect DNA methylation profiles of differentially expressed genes.

2.0 LITERATURE REVIEW

2.1 Strawberry

2.1.1 General introduction and Botany

Strawberry belongs to the Rosaceae family, consisting of over 3000 species divided into more than 100 genera (Shulaev et al., 2008). The Rosaceae encompasses a tremendous level of morphological and physiological variation including many herbs, shrubs and trees. Although leaf structure is usually alternate, the variation ranges from simple to palmately or pinnately compound. Members of this family possess stipules and their flowers are often attractive and bisexual. The sepals and petals are often pentamerous, and the number of stamens ranges from greater than 15 to fewer than ten. The carpel number ranges from one to many and the ovary structure varies by genus being either inferior or superior. The fruit can be a follicle, achene, pome, drupe, aggregate or accessory, and the endosperm is often lacking from the seed (Hummer and Janick, 2009).

Many of our most recognizable species belong to the Rosaceae including the economically important fruit and ornamental species. These include apples, pears, plums, peaches, cherries, apricots, almonds, brambles and strawberry as well as a number of ornamental species such as roses. The cultivated strawberry *Fragaria* x *ananassa* is grown extensively and well adapted for most temperate and some sub-tropical regions. From 2003 to 2013, the world strawberry production has increased from 5,041,331 tonnes to 7,739,622 tonnes (http://faostat3.fao.org/, 2016). In North America, strawberry production was 1,379,817 tonnes in 2013. Across the globe, 85% of strawberry production is distributed across the U.S., Mexico, Chile, Argentina, Brazil, Uruguay, Portugal, Spain, France, Italy, Morocco, Egypt, Turkey, China, Japan, and Australia (López-Aranda et al., 2011). Cultivated strawberry is one of the most important small berry fruits in Canada. In 2013, the harvested area of strawberry in Canada was 2477 ha and the production was 18,947 tonnes, mainly in Quebec, Ontario, British Columbia and Nova Scotia (http://faostat3.fao.org/, 2016).

A typical strawberry plant is composed of leaves, crowns, stolons, flowers and roots (**Figure 2.1**) (Childers, 1981). Strawberry has compound trifoliate leaves with long petioles. As a perennial plant, strawberry forms new roots at the base of crowns. The crowns, developing from axillary meristems, terminating with an inflorescence, can survive for several years. At

the upper side of the crown is a growing point. Stolons, a modified stem, can be developed from axillary buds, and this process is determined by genetic and environmental conditions. The propagation of strawberry can be sexual or vegetative, through either seeds or runners or branch crowns, with roots being formed at the node of new plants (Hollender et al., 2012).

2.1.2 Strawberry taxonomy and distribution

The *Fragaria* genus currently contains 23 species possessing different ploidy levels (Folta and Davis, 2006). The *Fragaria* genus has a ploidy series ranging from diploid to decaploid. Among twelve diploid species (2n=2x=14), the most common is the woodland strawberry *F. vesca* that is found throughout north temperate Europe, America, Asia and is also distributed in North Africa and South America. *F. vesca* contains four subspecies, the European subspecies *vesca* (Europe eastward to Siberia), three American subspecies *americana* (North America east of
the Rocky Mountains) , *bracteata* (North America west of the Rocky Mountains), and *californica* (restricted to California coast) (Stauct, 1988). Other diploid species show a narrower distribution, found only on a single continent, while others such as *F. viridis* occur widely in both Europe and Asia. Tetraploid species are found in Northern and Southwest China, Northeastern Asia and Himalayan region and hexploid species in Europe, eastward to Ural Mountains. Octoploid (2n=8x=56) *F. x ananassa,* originated from an accidental hybridization of two wild octoploid species, South American strawberries (*F. chiloensis*) and North American meadow strawberries (*F. virginiana*) (Folta and Davis, 2006).

2.1.3 Strawberry genetics and genomics

The octoploid strawberry species (*F.* × *ananassa*, *F. chiloensis*, and *F. virginiana*) genome composition was proposed to be highly dipoildized with the model AAA'A'BBB'B' (Bringhurst, 1990). This model indicated two genome types (A and B) with further differentiation into four subgenomes (A, A', B, and B') in the octoploid species. Phylogenetic analyses found *F. vesca* and *F. nubicola* were in the octoploids' ancestry (Potter et al., 2000). Identification of the relationship among the strawberry germplasms is important in strawberry breeding programs, using only morphological characteristics hasn't provided reliable information for the accurate identification of this relationship. To improve the reliability, molecular markers were widely used in genetic diversity studies and species genotyping. Randomly Amplified Polymorphic DNA (RAPD), AFLP markers have been used in

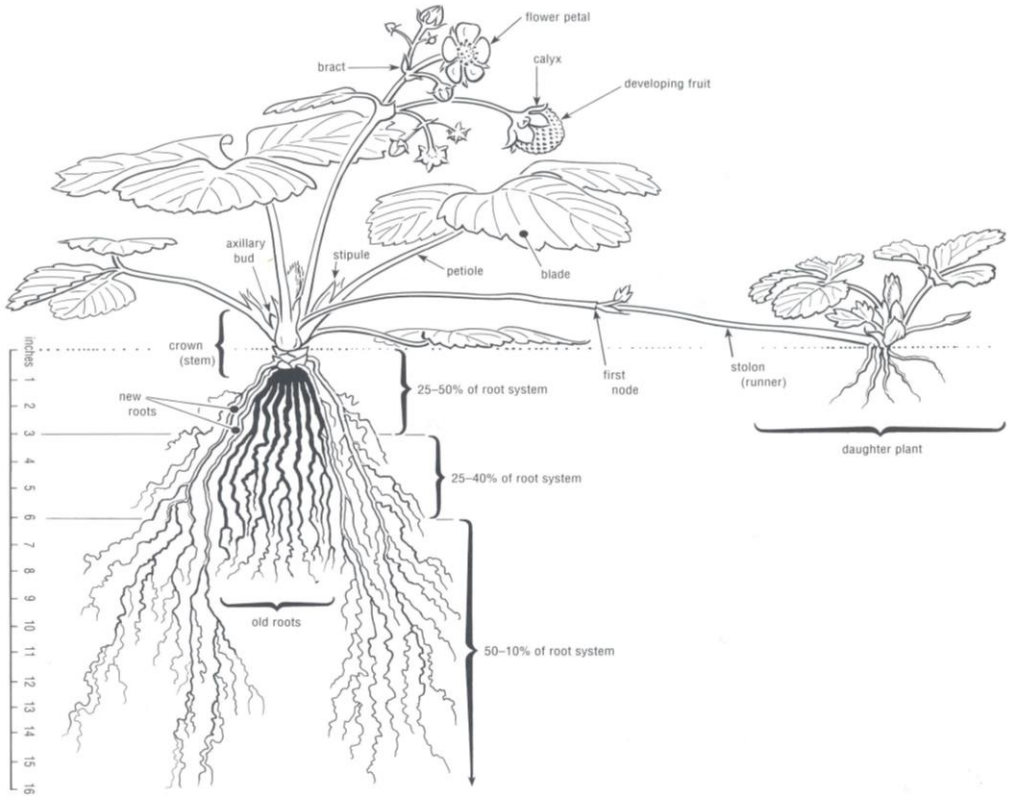


Figure 2.1 Structure of strawberry plant (from Integrated Pest Management for Strawberries, Publication 3351, Univ. Calif., 1994).

distinguishing and fingerprinting strawberry cultivars (Gidoni et al., 1994; Graham et al., 1996; Degani et al., 2001). The inter-simple sequence repeat (ISSR) marker was used in 30 geographically and genetically different strawberry cultivars classification, showing consistent results when compared to AFLP markers (Arnau et al., 2000).

The genetic and genomic complexities of octoploid strawberry limit its use in genomic research. However, diploid *F. vesca* plays an important role in connecting to other species (Bors, 2000). The genetic information of allopolyploid octaploid cultivars (*F.* × *ananassa*) indicates that it is descended from *F. vesca* (Senanayake and Bringhurst, 1967). *F. vesca* is self-compatible with many seeds on the surface of fruit called achenes. Therefore *F. vesca* is usually used as female parents to receive pollen from other self-incompatible species. The genome sequencing and assembly of *F. vesca* short reads derived from technologies Roche / 454, Illumina / Solexa and Life Technologies / SoliD platforms was recently completed (Shulaev et al., 2011). The first generation annotation (version 1.1) was generated and can be found in the Genome Database for Rosaceae (GDR, http:/www.rosaceae.org) using GeneMark, resulting in 34809 predicted genes. Functional annotation identified around 25050 preliminary annotated genes (Shulaev et al., 2011; Jung et al., 2014). The revised annotation was generated using MAKER2 based on five stages with five fruit tissue types in the early-stage fruit development, with more predicted protein coding genes compared to the previous annotation (Darwish et al., 2015). Small genome size (240 Mb) is an attractive feature for *F. vesca*. The basic set chromosome number x=7 of *F. vesca* makes the genome size smaller than many other Rosaceae members: peach (270 Mb), raspberry (294 Mb), pear (539 Mb), and apple (750 Mb) (Folta and Davis, 2006). These strawberry genomic resources combined with the next generation sequencing technology and bioinformatics tools will make it possible to comprehensively study gene expression and DNA methylation profiles at whole genome level in *F. vesca*.

2.1.4 *Fragaria vesca*

The diploid woodland strawberry *F.vesca* is regarded as a valuable model plant in the Rosaceae family for genetic and linkage map studies (Davis and Yu, 1997; Sargent et al., 2009). The genetic map between diploid and octoploid *Fragaria* species showed a high level of colinearity in the seven expected homologous groups (Rousseau-Gueutin et al., 2008). A further advantage is the ease of transformation using *Agrobacterium* and rapid regeneration from tissue culture (Oosumi et al., 2006; Rabinowicz et al., 2008). In addition, the small

genome size and simple ploidy level encouraged the adoption of *F. vesca* for genome sequencing. The completion of this achievement opened the door to post-genomics analyses in the Rosaceae. These properties of *F. vesca* make it an ideal model system for genomic and gene function study with a direct relationship to more complex polyploidy species as well as other members of the Rosaceae family. Three *F. vesca* inbred lines Ruegen F7-4, 5AF7, and Hawaii 4 have been widely used for isolating genes and characterizing gene functions over the past few years (Mouhu et al., 2009; Hollender et al., 2012). The Ruegen F7-4 originates from Germany, and is an inbred line of Ruegen that possesses a red berry. The 5AF7 is an inbred line of Yellow Wonder and possesses a white berry. The Hawaii 4 originates from the wild in Hawaii, it bears white-yellow fruits and has runners while the other two ecotypes lack the ability to produce runners. The genome sequencing of Hawaii 4 was completed using the fourth generation inbred line 'H4×4' (Shulaev et al., 2011).

2.1.5 Strawberry plant growth and development

The growth cycle of Rosaceae plants varies depending on different plant types of herbs, shrubs, and trees. The juvenile stage also varies from several years (fruit trees such as apple) to months (such as strawberry) (Visser, 1964). Perennial strawberry has repeated annual growth cycles of vegetative and reproductive stages, however, the June-bearing (seasonal, short day, hereon referred to as short day) strawberry and everbearing (perpetual, day neutral, hereon referred to as day neutral) strawberry possess different yearly growth cycles as their growth and development are affected by photoperiod and temperature (Brown and Wareing, 1965). In the short day (SD) strawberry, flower initiation begins as the days get shorter with the competence of stolon formation and low temperature (Heide, 1977). The branch crowns are developed from axillary buds when stolon production ceased. After winter semi-dormancy, in spring the plants resume growth and the inflorescences emerge with subsequent fruit development (Darrow, 1936). In day neutral (DN) strawberry, the growth and development is different from the SD strawberry. The DN strawberry has a short juvenile phase, and both floral initiation and flowering take place in the first growing season. Continuous flowering and fruiting starts from early summer and lasts until late autumn (Brown and Wareing, 1965; Kurokura et al., 2013).

2.1.5.1 Strawberry propagation and growth

Strawberry propagation is botanically through either seeds or clonal daughter plants, however,

daughter plants are typically commercially propagated. Strawberry achenes exhibit relatively low germination rate because of the tegument dormancy due to the hard texture pericarp (Galvão et al., 2014). Scarification can be used to overcome dormancy by breaking the seed coat, increasing water and gas permeability (Mayer and Poljakoff-Mayber, 1982). In strawberry, besides scarification of achenes (Miller et al., 1992), studies reported using other methods such as pre-chilling treatment, light exposure and chemical pretreatment to break dormancy and promote germination (Wilson et al., 1973; Miller et al., 1992; Galvão et al., 2014). Germination starts with the uptake of water and the visible sign of ending is the emergence of the radicle from its surrounding seed coat. The shoot, containing the cotyledons, emerges in the following phase after radicle emergence (Bewley, 1997). The cotyledons expand and support the growing radicle and stem, and at the cotyledon stage the plants can be transplanted into soil.

In commercial strawberry production, in order to obtain genetically uniform seedlings and have earlier fruit bearing time, propagation is often clonal through daughter plants. During vegetative reproduction, the axillary bud, which will later develop into genetically identical daughter plants on modified stems (stolons), is formed at the terminal of the first internode. Stolons connect mother plants forming a colony of genetically identical individuals. The stolon connection allows the transportation of nutrition from the mother to daughter plants to assist in establishment of the newly forming clone (Childers, 1981). The earlier stolon daughter plants grown from mother plants were selected for propagation as they exhibit higher survival rate. The first four to five leaves in strawberry are simple leaves, and then the leaves grow into compound trifoliate leaves (Chatterjee et al., 2011). The leaves produced later are smaller and when the leaf numbers is around six, the inflorescences start to emergence. Sexual reproduction occurs in a mature inflorescence and includes the primary flower. This structure which terminates the growth axis is followed by secondary, tertiary and quaternary, and even quinternary flowers. Each flower possesses five petals, five sepals, and many stamens and pistils (Shoemaker, 1978). Pollination and fertilization of strawberry can be completed within the same flower or among the flowers of the same or different cultivars.

2.1.5.2 Regulation of floral and fruit development

The shoot apical meristem (SAM) is a collection of cells on the tip of the growing shoot, and generates all aboveground plant organs. In strawberry, the inflorescences are formed from the

apical meristem of the branching crowns, and therefore the formation of crown branches and axillary buds are the prerequisite for satisfactory flowering. The crown meristems can be classified into primary, secondary, or tertiary meristems. The primary meristem is described as the apical meristem from the main crown. The axillary meristem from the main crown and branch crowns are referred to as the secondary and tertiary meristem respectively (Hytönen et al., 2004). After the transition from vegetative growth to reproductive growth, the physiological and morphological changes occur at the apical meristem, followed by the differentiation and development of floral organs and flowering. The anatomy and morphology changes during the phases of flower initiation and differentiation was described in previous research in cultivated octoploid strawberry (*Fragaria* x *ananassa*) (Jahn and Dana, 1970; Taylor et al., 1997). A recent study in *F. vesca* flower development indicated the similarity of floral structure and floral organ development compared to *F.* x *ananassa* except the flower size is smaller and the number of stamens is less in *F. vesca* (Hollender et al., 2012). The development of *F. vesca* flower was divided into 13 stages, with three separate segments, early flower development, reproductive organ initiation and development, and floral organ differentiation based on a study using scanning electron microscopy and stereomicroscopy (Hollender et al., 2012). Flowering is defined in this thesis to occur when it reaches stage 12 with fully developed organs and visible white petals.

F. vesca is self-compatible and achenes are produced on the surface of soft fruit, botanically named as aggregate fruit. The berry is a collection of many small achenes (approximately 1mm) in spirally arranged rows and the enlarged receptacle which is edible. The mature achene contains a hard texture pericarp and shows dormancy. Fruit development and ripening of strawberry is relatively fast among the Rosaceae. It usually takes a month for the strawberry fruit to mature after pollination and fertilization. Generally speaking, in most plants, fruit development consists of three distinct stages: Stage 1 is ovary development, fertilization and fruit set; Stage 2 is fruit growth due to cell division, accompanied by seed formation and early embryo development; Stage 3 is fruit enlargement by cell expansion and embryo maturation phase (Gillaspy et al., 1993). The fruit development is composed of six stages including small green, medium green, large green, white, turning and red (Fait et al., 2008). For *F. vesca* Hawaii 4 the ripen fruit color is white-yellow. In strawberry the achenes produce auxin which is essential for receptacle enlargement. Auxin level peaks during the first stage and it can be translocated basipetally through the phloem (Dreher and Poovaiah, 1982; Friml, 2003). The ripening or softening of the receptacle is related to the decrease of

auxin levels (Archbold and Dennis Jr, 1985). Failure of berry development was reported in fruit without achenes on the surface of the receptacle but exogenously applied auxin can replace achenes to maintain the growth of fruit (Nitsch, 1970). The fruits tend to have longer length than diameter, the shape and size of berry are determined by flower bud position, cell division and expansion (Darrow, 1966). Unlike other Rosaceae crops such as apple and peach, which are climacteric, the strawberry fruit becomes ripe without ethylene production and respiration increases (Perkins-Veazie, 2010).

2.2 Epigenetic modifications in plants

2.2.1 Epigenetics definition and mechanisms

Epigenetics is defined as the heritable changes in phenotype or gene expression by mechanisms that do not alter the primary DNA sequence (Riggs and Porter, 1996). These changes are caused by the presence of functionally relevant modifications to the chromosome that add onto the DNA sequence (Rival et al., 2010). However, the term epigenetics started out with a wider definition and has been repeatedly refined over the last seventy years. It was first coined by Waddington to describe cellular differentiation from embryonic stem cells throughout ontogeny and represented the interaction between genes and their environment to ultimately induce a phenotype (Waddington, 1942). More recently, the term epigenetics was used to describe the regulation of gene activity in time and space and their effect on the development of complex organisms (Holliday, 1994). It is now accepted that the definition has further narrowed to include the study of stable heritable alternations in gene expression resulting from changes other than those encoded in the primary DNA sequence that are transmitted through mitosis and/or meiosis (Morris, 2001). Several biochemical mechanisms underlying epigenetic phenomena have been reported including the systems of DNA methylation, post-translational modification of histone proteins and non-coding RNA molecules (Jablonka and Raz, 2009). These mechanisms can interact to organize the structure and configuration of the DNA molecule and provide additional layers of transcriptional machinery that can result in the activation or repression of gene expression (Saze, 2008; Feng et al., 2010).

2.2.1.1 DNA methylation

DNA methylation is recognized as a well-studied epigenetic mark in plants and mammals (Bird, 2002; Reinders et al., 2009). In eukaryotes, DNA methylation is referred to as the

presence of a methyl group at position five of a cytosine pyrimidine ring, a process that is mediated by DNA methyltransferases which catalyze methylation of cytosine. In plants, the location of methylated cytosine are found in both symmetric (CG and CHG) and asymmetric (CHH [where H=A, C, or T]) contexts. This contrasts with DNA methylation in mammals that are largely restricted to the symmetrical CG dinucleotide context (Oakeley and Jost, 1996). In *Arabidopsis*, the methylation at symmetric CG and CHG sites can be maintained during DNA replication through the action of DNA maintenance enzymes such as DNA METHYLTRANSFERASE 1 (MET1, homologue of the mammalian Dnmt1) and CHROMOMETHYLASE 3 (CMT3) (Lindroth et al., 2001; Kankel et al., 2003).Whereas the *de novo* methylation enzymes including DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) are recruited by the RNA polymerase Pol V, a silencing complex ARGONAUTE 4, and 24-nt small interfering RNAs (siRNAs) in a sequence-specific manner to establish methylation at CHH sites (Goll and Bestor, 2005; Law and Jacobsen, 2010). This RNA-directed DNA methylation (RdDM) was first explained in tobacco plants (Wassenegger et al., 1994). Recently reports in *Arabidopsis thaliana* and *Brassica oleracea* indicated lower DNA methylation levels in CHH than in the CG and CHG methylation sites. In *Arabidopsis*, the methylation levels of CG, CHG and CHH was 24%, 6.7% and 1.7%, respectively. In *Brassica oleracea*, the methylation levels of CG, CHG and CHH was 54.9%, 9.4% and 2.4%, respectively (Cokus et al., 2008; Parkin et al., 2014).

Major differences in cytosine methylation patterns throughout plant genomes distinguishes gene-rich regions from repeat-rich regions but variation in methylation among genes and their promoters has been shown to control expression (Bender, 2004). Altering the constraints provided by DNA methylation to gene expression might manifest as changes in phenotypes. Additionally, alterations in DNA methylation patterns can be indirectly mutagenic, affecting the higher-order chromosome conformation leading to transposon activation (Wolffe and Matzke, 1999; Martienssen and Colot, 2001). Epigenetic phenomena have been described long before the mechanisms of their transmission were known. In 1809, the French naturalist Lamarck proposed the law of inheritance of acquired traits. It was often noted as oddities that appeared to defy the laws of Mendelian inheritance by geneticists after 1930s (Burkhardt, 1979; Sano, 2010). The agronomist Lysenko in the Soviet Union also supported the law of inheritance of acquired characters with a large number of experiments conducted in winter and spring wheat during 1930s and 1960s (Steele et al., 1999). A classic epigenetic

phenomenon is paramutation which was first reported in leaves and petal rogue phenotypes in peas (Bateson and Pellew, 1915). The study of paramutation in maize was begun in the seed pigmentation phenotype determined by the r1 (red 1) locus and later found at other loci (booster 1, plant color 1, and pericarp color 1). These loci are responsible for tissue pigmentation which is easy to visualize (Chandler and Stam, 2004). The heritable phenotypic changes resulted from gene expression changes, caused by epigenetic information transferred from one allele of a gene to another allele with trans interactions (Stam, 2009). It has been established that epigenetic changes in DNA methylation patterns are responsible for the changes from bilateral asymmetry to radial symmetry (peloric mutants) of the floral structure in toadflax (*Linaria vulgaris*). Lcyc is a gene in toadflax responsible for bilateral asymmetry floral structure. Peloric mutants in toadflax show a lack of Lcyc gene expression. This is because the Lcyc gene is highly methylated and transcriptionally inactive in peloric mutants while in wild type plants it is partially demethylated leading to adequate levels of expression (Cubas et al., 1999).

2.2.1.2 Histone modification

In eukaryotes, chromatin consists of nucleosome units formed by DNA, histones and non-histone proteins. A nucleosome unit contains an octamer of eight histones, consisting of two sets of H2A, H2B, H3 and H4 histones, wrapped by 147 base pairs of DNA (Karlić et al., 2010). Heterochromatin which is usually associated with low levels of gene expression can be divided into constitutive heterochromatin and facultative heterochromatin. Constitutive heterochromatin is still condensed and transcriptionally silent throughout the whole cell cycle. By contrast facultative heterochromatin in the condensed state can be regulated by cellular signals (Grewal and Jia, 2007). Histone modification represents post translational modifications of the N-terminal and C-terminal regions of core histone proteins. The various modifications of histones include acetylation, methylation, phosphorylation, ubiquitination and ADP-ribosylation (Sridhar et al., 2007; Zhang et al., 2007). Histone acetylation is intensively studied among these modifications, especially H4 and H3, and is regulated by histone acetyltransferase (HAT) and histone deacetylase (HDAC). Histone acetylation loosens the affinity of histone protein H4 with DNA, which cause transcriptionally active state of chromatin, meanwhile deacetylation of histone tend to act as a transcriptional repressive chromatin modification (Rakyan et al., 2001). These covalent modifications of the histone complex and dynamic changes in chromatin structure are related to the gene expression

changes in plants.

Three types of histone methylation in plants were well-studied, including methylation of lysine 4 on histone H3 (H3K4), lysine 9 on histone H3 (H3K9), and lysine 27 on histone H3 (H3K27) (Feng and Jacobsen, 2011). Lysines are able to accept up to three methyl groups to form mono-methylation, di-methylation, and tri-methylation. In the winter-annual *Arabidopsis*, a period of cold temperature is required in order to flower, a process called vernalization (Chouard, 1960). The mechanism underlying this phenomenon was the downregulation of the flowering repressor gene *FLOWERING LOCUS C* (*FLC*) to promote flowering (Koornneef et al., 1994; Michaels and Amasino, 1999). Studies indicated histone modification was involved in this regulation, for example the enrichments of H3K9 dimethylation and H3K27 dimethylation (Bastow et al., 2004; Sung and Amasino, 2004). Additionally, the expression of four genes induced by drought stress was related to the increasing H3K4 trimethylation and H3K9 acetylation in *Arabidopsis* (Kim et al., 2008).

2.2.1.3 Interaction of DNA methylation and histone modification

Epigenetic regulation of gene expression is a complex process with different factors involved as mentioned, as well as the interaction among these factors. The state of DNA methylation was correlated with the patterns of histone modifications that can regulate gene transcription level (Lippman et al., 2003). In *Arabidopsis*, some genes modified with H3K27 trimethylation usually lack DNA methylation, showing histone modification conflicts with DNA methylation (Zhang et al., 2007). In a salt stress study of *Arabidopsis*, the reduced gene expression and hypermethylation in promoters were correlated with increased of H3K9 dimethylation and decreased of H3K9 acetylation of the progeny (Bilichak et al., 2012). Altered DNA methylation and histone methylation and acetylation were observed in the artificially selected lines of high energy use efficiencies in *Brassica*. Compared to the control, the level of methylation of H3K9, H3K27 and acetylation of H4K8, H4K12, H4K16 was lower than selected lines (Hauben et al., 2009).

2.2.2 5-azacytidine as a mutagen to alter DNA methylation

Besides the naturally occurring epimutations, established patterns of DNA methylation can be stochastically changed through the application of potent pharmacological agents such as 5-azacytidine (5-azaC). This compound is an analogue of cytidine possessing nitrogen rather

than a carbon atom at the 5-position of the pyrimidine ring. During DNA replication 5-azaC is randomly incorporated into the genome resulting in demethylation at cytosine positions (Christman, 2002; Stresemann, 2008). Studies altering genome-wide DNA methylation patterns have demonstrated exposure to exogenous 5-azaC in plants can increase phenotypic trait variation, some of which are heritable. Strategies exploiting this phenomenon are being applied in a range of plant species where it has been demonstrated that dwarfism in *Oryza sativa* (Sano et al., 1990), early flowering in *Arabidopsis* (Burn et al., 1993), *Linum usitatissimum* (Fieldes, 1994b; Fieldes and Amyot, 1999; Fieldes et al., 2005), and *Solanum ruiz-lealii* (Marfil et al., 2012), as well as abnormal flower development, and leaf morphology *Solanum ruiz-lealii* (Marfil et al., 2009; Marfil et al., 2012) can be identified and selected. 5-azadeoxycytidine (5-azadC), the deoxy version of 5-azaC , is another analogue of cytosine that also inhibits DNA methylation (Jüttermann et al., 1994) and can result in novel phenotypes due to hypomethylation in *Oryza sativa* (Akimoto et al., 2007). Based on our knowledge the mutagenic effects and chromosomal rearrangement of 5-azaC have been studied in mammalian (Viegas-Péquignot and Dutrillaux, 1976; Broday et al., 1999). However, later studies indicated the genomic integrity loss is not obligatory, as mentioned in this review (Bird, 2002). Furthermore, study published in the same year verified that 5-azacytidine has little ability to change primary DNA sequence when it works as an efficient DNA methylation inhibitor (Holliday and Ho, 2002). However, few reports have comprehensively verified no mutagenic effect induced by 5-azaC in plants, although most studies treat it as an epimutagen reactivating genes without changing primary DNA sequencing.

2.2.3 DNA methylation detection

Approaches of DNA methylation analysis have been reviewed previously including global methylation and locus specific methylation (Oakeley, 1999; Shen and Waterland, 2007; Beck and Rakyan, 2008). From cytosine position nonspecific global DNA methylation to high throughput single base resolution methylation profiling, many technologies such as High Performance Liquid Chromatography (HPLC) (Kuo et al., 1980), Methylation Sensitive Amplified Polymorphisms (MSAP) (McClelland et al., 1994), Microarray (Gitan et al., 2002), Methylated DNA immunoprecipitation (MeDIP) (Weber et al., 2005), Bisulfite treatment coupled with next generation sequencing (Krueger et al., 2012) have been emerged and used for DNA methylation analysis. In strawberry, previous studies reported using HPLC to detect DNA methylation (Zhang et al., 2012) or applying MSAP method with isoschizomers to

digest genomic DNA (Hao et al., 2002; Martelli et al., 2008). However, comprehensive studies of the strawberry methylome have not been widely reported.

2.2.3.1 Methylation Sensitive Amplified Polymorphisms

Methylation Sensitive Amplified Polymorphisms (MSAP), also referred to as MS-AFLP (Methylation Sensitive AFLP) is a modified method used to investigate the extent of cytosine methylation of the 5'-CCGG-3' site based on amplified fragment length polymorphism (AFLP) (Xiong et al., 1999). This technique was first reported in a study of DNA methylation in dimorphic fungi (Reyna-López et al., 1997). MSAP is able to detect differential methylation patterns by using isoschizomers *Hpa* II and *Msp* I to replace *Mse* I in the AFLP method as the frequent cutting enzymes. Both enzymes *Hpa* II and *Msp* I can recognize the same restriction (5'-CCGG-3') site but cleavage of DNA is dependent on the methylation context of each cytosine base in the recognition site (McClelland et al., 1994). In this way, MSAP can be used as a method to detect DNA methylation changes. The cytosine methylation profiles of 5'-CCGG-3' contexts in flax (Brown et al., 2008), rice (Kou et al., 2011), pepper (Portis et al., 2004), rose (Xu et al., 2004) and many other crops have been detected using this technique. However, the drawbacks of this technique are that it can only recognize 5'-CCGG-3' across the genome and not all theoretical profiles of double stranded 5'-CCGG-3' can be identified. Further, different methylation profiles may correspond to the same digestion pattern of *Hpa* II and *Msp* I (Fulneček and Kovařík, 2014; Rico et al., 2014).

2.2.3.2 Methylated DNA immunoprecipitation

MeDIP technique is an efficient method to detect DNA methylation at both whole genome and loci specific level with limited amounts of DNA samples (200 ng) (Weber et al., 2005; Thu et al., 2009). In this method, antibodies are used to identify methyl-cytosines from sheared genomic DNA but not limited to specific restriction sites, and this can be coupled with high-resolution microarrays or next generation sequencing to reflect methylated sequences, especially CG rich sequences. Compared to MSAP, the MeDIP is a more unbiased method to detect methylated DNA even though the resolution of this method is a few hundred base pairs or less. The first methylome profile in plants was reported using this method in *Arabidopsis* (Zhang et al., 2006; Zilberman et al., 2007). The appropriate unmethylated control was important as the immunoprecipitation was affected by the factors such as CG density and repetitive elements (Thu et al., 2009).

2.2.3.3 Sequencing-based methylation profiling

Bisulfite conversion of genomic DNA enables the detection of all methylated cytosines by treating the genomic DNA with sodium bisulfite (Susan et al., 1994). This treatment causes deamination of unmethylated cytosine residues in fragmented DNA, turning cytosine into uracil. After subsequent polymerase chain reaction (PCR) amplification uracil was converted into thymine. While the methylated cytosine is not converted, the DNA methylation profiles can be analyzed by comparing Sanger sequencing information of the PCR product to the unconverted reference (Clark et al., 2006).

Recently, next generation sequencing platforms with higher-throughput and more advanced computational processing of short reads sequencing make it possible for single base resolution profiling of DNA methylation at the whole genome level (Mardis, 2008; Xi and Li, 2009). The bisulfite conversion of genomic DNA combined with next generation sequencing approach was termed BS-Seq, providing more sensitive measurement of cytosine methylation on genome-wide scale compared to previously reported microarray based study (Zhang et al., 2006; Zilberman et al., 2007). The BS-Seq is known as a gold standard method with single base resolution to measure whole genome DNA methylation. The epigenome of *Arabidopsis thaliana* based on BS-Seq was independently reported by Lister et al., (2008) and Cokus et al., (2008) studies. In the study by Cokus and colleagues, a cluster of five methylated CG sites in FLOWERING WAGENIGEN (FWA) locus were detected. FWA is a flowering repressor and, ectopic expression of FWA induced late flowering in *Arabidopsis* (Kakutani, 1997). Cokus et al. finding is complementary to a previous study (Zhang et al., 2006). Additionally, BS-Seq was found to be more accurate in detecting promoter methylation of genes compared to previous microarray method (Cokus et al., 2008). Therefore, compared with other methods to test DNA methylation, single base resolution mapping of DNA methylation can answer questions about methylation profiling in different regions of the genome, including gene body, upstream, downstream, transcriptional termination regions, transposable element (TE). It is a most comprehensive way to study DNA methylation.

2.3 Epigenetic inheritance and selection of epigenetic variants

2.3.1 Epigenetic inheritance through meiosis and mitosis

Epigenetic inheritance is the transmission of acquired traits from parents to subsequent generations without changes in DNA sequence. The inheritance of epigenetic variation in

plants where marks are transmitted through mitosis or meiosis continues to remain controversial although there is a growing body of evidence demonstrating its efficacy (Kakutani et al., 1999; Heard and Martienssen, 2014). In order to remove the genetic contributions to phenotypic traits, genetically identical or doubled haploid lines (Hauben et al., 2009), epigenetic recombinant inbred lines (epiRILs) (Johannes et al., 2009; Reinders et al., 2009; Zhang et al., 2013), or clonal plant lines (Gao et al., 2010; Verhoeven et al., 2010; Raj et al., 2011) were used to study the epigenetic effect. The epiRILs were derived from two parents with almost identical DNA sequence but divergent DNA methylation profiles, as one parent is homozygous ddm1 or met1 mutant lack of DNA methyltransferase. In the model plant *Arabidopsis thaliana* epigenetic variation and high heritability was observed in complex quantitative traits such as flowering time, plant height and plant growth in epiRILs (Johannes et al., 2009; Reinders et al., 2009). Subsequent field and greenhouse experiment of phenotyping the epiRILs further indicated the heritable variation for these complex traits in response to drought and nutrient conditions (Roux et al., 2011; Zhang et al., 2013).

In addition to working with model species, examples of naturally occurring non-Mendelian inherited traits are known, for example, spontaneous epimutations resulting from hypermethylation in the promoter of an SBP-box transcription factor have been shown to inhibit tomato fruit ripening (Manning et al., 2006). The heritable epigenetic changes were caused by single locus cytosine methylation variations (epialleles). Moreover, epigenetic inheritance of dwarfism in rice over nine generations induced by 5-azadC was reported (Akimoto et al., 2007). Similarly, the transgenerational inheritance of seed size, plant stature and floral morphology have been studied in *Brassica* derived from 5-azaC treatment (Amoah et al., 2012).

The mitotic inheritance is referred to as the propagation of epigenetic modifications through mitotic cell division (Martin and Zhang, 2007). The asexual propagation leads to clonal persistence of the induced and selected phenotypic variations. In vernalization experiments in *Arabidopsis*, the floral repressor *FLC* is silenced after exposure to cold temperature through changes of the epigenetic modifications (Amasino and Michaels, 2010). This modification status can be stably transmitted mitotically to induce flowering, and the *FLC* silencing status remained the same in the following generated plant organs. However, the epigenetic modifications of *FLC* were reset in the next generation progeny of vernalized plants through sexual propagation (Amasino and Michaels, 2010). Although we reviewed the faithful

transgenerational inheritance of epigenetic variations, some epigenetic marks can be reset at meiosis and this probably explains why the phenotypic trait induced by epigenetic changes disappeared in the following generations (Hauben et al., 2009). In the stress-induced *Arabidopsis* lines possessing epigenetic memory, the modified epigenome can be transmitted to the next generation, but it disappeared after two generations without stress treatment (Suter and Widmer, 2013).

2.3.2 Selection of epigenetic variants

DNA methylation alterations can create stable meiotically heritable epialleles through both naturally occurring and artificially generated methods (Salmon et al., 2008; Schmitz et al., 2011; Amoah et al., 2012). The isogenic *Arabidopsis* plants BS-Seq study found the frequency of detected epialleles changes in each generation was higher than the alleles in genetic mutation (Ossowski et al., 2010; Hirsch et al., 2012). This suggested that heritable epialleles may be possible under selection if they cause a phenotype. The generation of epiallelic variation and the resultant phenotypic diversity that allowed selection was demonstrated in *Brassica rapa* inbred lines (Amoah et al., 2012). The reduced plant stature and small seed size (E2 generation) were heritably derived through selfed-seeds selected from the E1 generation following 5-azaC treatment (E0 generation). In *Arabidopsis*, quantitative genetics methods were used to study the heritability of ecologically important traits such as flowering time, plant height, total biomass and root:shoot ratio in different environmental conditions with selection applied. The result indicated the action of selection on the plasticities of certain traits (Zhang et al., 2013). Another example of successful artificial selection of epimutant lines was reported in *Brassica napus* (Hauben et al., 2009). The selected lines were self-fertilized and progenies were then generated with distinct energy use efficiency after three to five rounds of selection. The selected inherited lines were used to study the epigenetic status and found hypomethylation exists in these lines especially in the lines with low respiration, and low histone methylation and acetylation (Hauben et al., 2009; Verkest et al., 2015). However, the study also indicated the epialleles may fail to transmit into the following generations when the selection pressure is not applied in a population (Hirsch et al., 2012).

2.4 Flowering and genes involved in the flowering pathway

2.4.1 Flowering time study in *Arabidopsis thaliana*

Flowering is a multi-gene controlled process and the initiation of flowering is determined by both environmental (external) and endogenous (internal) factors. This complex process has been most extensively studied using the model plant, *Arabidopsis*. The environmental cues mainly include photoperiod and temperature (vernalization). The endogenous pathways (gibberellin, autonomous pathways) are independent of environment and related to the development stage of plant (**Figure 2.2**). In the juvenile phase, the plants are not competent to flower even though the environmental inputs are inductive for flowering. However, after transition from juvenile to adult phase the appropriate environmental stimulus will promote flowering (Greenup et al., 2009; Kim et al., 2009; Amasino, 2010). Flowering time is an important phenotype as it is correlated to time of seed set, and in nature it is very important for the plant to complete its life cycle before frost from the perspective of survival. Flowering time easily undergoes natural or artificial selection which may be epigenetically controlled (Rapp and Wendel, 2005).

Photoperiod is a major determinant to regulate the transition from vegetative to reproductive growth in plants. Based on the response to photoperiod, plants can be classified into long-day (LD) in which the floral transition occurs when the day length exceeds a critical length, short-day (SD) in which flowering occurs when the day length is less than a critical length, and day-neutral (DN) plants in which flowering is induced regardless of photoperiod. LD plants and SD plants can be further divided into facultative and obligate types. For example, *Arabidopsis* is a facultative LD plant, flowering earlier under long days but flowering more slowly under short days, while oat is an obligate LD plant, flowering will not occur unless under the long day photoperiod.

Photoperiodism is associated with circadian rhythms and endogenous circadian rhythms are controlled by the plant circadian clock (Dunlap, 1999). The circadian clock comprises three molecular feedback loops including the central loop, morning loop, and evening loop (Fornara et al., 2010). Many *Arabidopsis* genes and transcription factors are involved in the feedback circuits including two MYB transcription factors *LATE ELONGATED HYPOCOTYL* (*LHY*) and *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*), which repress *TIMING OF CAB EXPRESSION1* (*TOC1*) while LHY / CCA1 were repressed by *PSEUDO-RESPONSE REGULATOR7* / *9* (*PRR7* / *PRR9*) in the morning loop. *CONSTANS* (*CO*), a B-box-containing protein, is the target of circadian clock genes *GIGANTEA* (*GI*) and F-box ubiquitin ligases (FKF1), which promote the expression of *CO* by down-regulating its

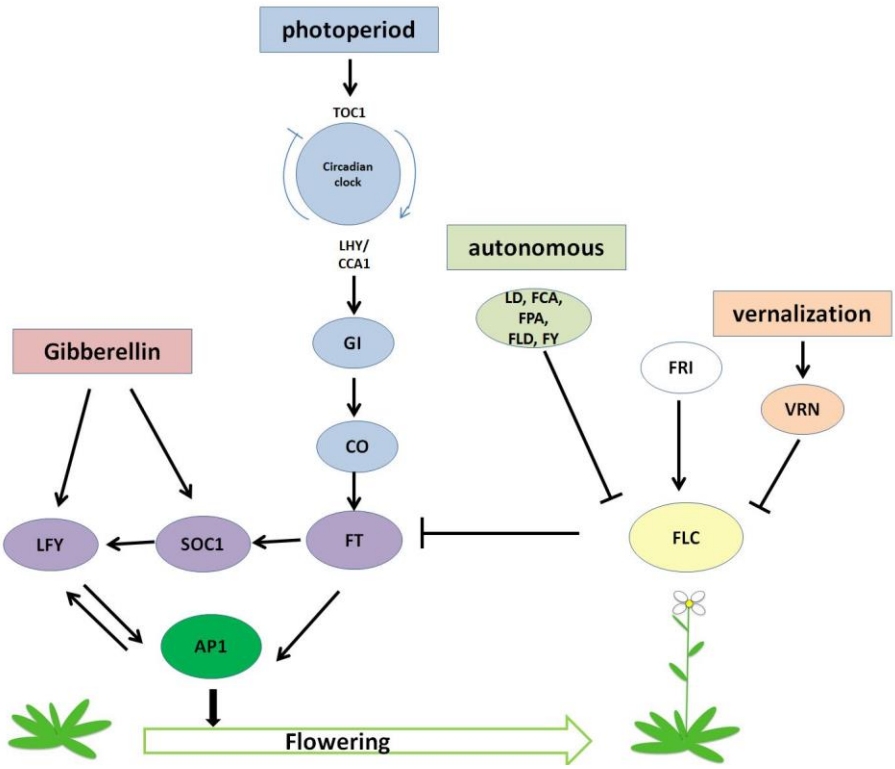


Figure 2.2 Flowering time pathway in *Arabidopsis*.

This figure represents the relationships among different factors (gibberellin, autonomous, photoperiod, and vernalization) in flowering time pathways in the model plant *Arabidopsis*. Arrows mean positive regulation and bars negative regulation. For full gene names see list of abbreviations.

transcriptional repressor *CYCLING DOF FACTORs* (*CDFs*) (Fornara et al., 2010; Johansson and Staiger, 2015). *CO* is also the key gene in the *Arabidopsis* photoperiod pathway to promote flowering under long days by activating the transcription of the downstream gene *FLOWERING LOCUS T* (*FT*) (Kardailsky et al., 1999). At first FT mRNA was reported as the florigen signal, and in a later study the FT protein was referred to as florigen (Huang et al., 2005; Jaeger and Wigge, 2007). Since flowering occurs in the SAM, the photoperiod and the above response is perceived in the leaves. The mobile signal FT protein needs to move from the leaves to the SAM through phloem translocation and activates other flowering integrator genes *SUPPRESSER OF OVEREXPRESSION OF CO1* (*SOC1*) and *LEAFY* (*LFY*) (Moon et al., 2005; Corbesier et al., 2007). These key genes are the common downstream target genes connecting all flowering pathways to form an integrated network. During the transition to flowering, the fate of the meristem changes from the vegetative to the floral phase and stops leaf production. During this transition, the increased level of *SOC1* activates the expression of *LFY*, and *LFY* further activates the expression of *APETALA1* (*AP1*) in reproductive meristems (Hempel and Feldman, 1994; Levy and Dean, 1998; Yu and Ma, 2001). These genes promote flowering while some genes repress flowering such as *TERMINAL FLOWER 1* (*TFL1*). *TFL1* is a flowering repressor in meristematic responses and developmental pathways in *Arabidopsis* and represses floral meristem identity genes *AP1* and *LFY* expression in shoot and inflorescence meristems (Simon et al., 1996).

Many *Arabidopsis* types with a winter annual growth habit need to experience a period of low temperature to become competent to flower, called vernalization. FLC, a MADS box transcription factor, represses flowering through repressing *FT* and *SOC1* (Michaels and Amasino, 1999). *FRIGIDA* (*FRI*) is a gene activating the expression of *FLC*, most early flowering Arabidopsis varieties contain the mutant of *FRI* which cannot promote the expression of *FLC* (Mouradov et al., 2002). *FLC* and *FRI* are the main genes involved in the vernalization pathway. The plant hormone Gibberellin (GA) accelerates flowering in *Arabidopsis*. The effects of mutations *ga1*, *ga4* and *ga5* in controlling flowering time have been studied (Wilson et al., 1992; Mouradov et al., 2002). Studies indicated GA promotes flowering by upregulating the expression of *LFY* and resulting in earlier *FT* transcription under SD conditions in *Arabidopsis* (Gocal et al., 2001).

2.4.2 Flowering time study in *Fragaria vesca*

The molecular mechanisms of flowering in the annual plant *Arabidopsis thaliana* have been comprehensively understood. However, the molecular control of flowering time in perennials has only been recently reported. In perennials, plants experience repeated seasonal cycles of vegetative and flowering stages that are controlled by seasonal flowering genes (Wang et al., 2009). In strawberry the flowering transition was influenced by both temperature and photoperiod (Darnell et al., 2003; Tanino and Wang, 2008). Recently, diploid *F. vesca* has been used as a model system to study flowering time in Rosaceae (Mouhu et al., 2009; Mouhu et al., 2013). In *F. vesca* there are SD and DN flowering habits. Classic genetic studies indicated the perpetual flowering genotype is caused by floral repressor *SEASONAL FLOWERING LOCUS* (*SFL*) recessive alleles (Brown and Wareing, 1965; Albani et al., 2004). It demonstrated the seasonal genotypes of strawberry flower faster under short days and cool temperature (Sønsteby and Heide, 2008). The June-bearing strawberry cultivars are normally facultative SD plants (Hartmann, 1947). However, there is some controversy in the photoperiod responses of perpetual strawberry. Previous studies in the 1980s indicated the perpetual genotypes are day-neutral (Smeets, 1980; Durner, 1984). However, later studies showed that LD and higher temperature can accelerate the induction of flowering in perpetual flowering accessions (Sønsteby and Heide, 2007, 2008). Most recently, the perpetual strawberry was found to be continuously flowering and insensitive to photoperiod (day-neutral) after floral induction, indicating the day-neutral response to photoperiod in later stages of flowering
(Koskela et al., 2012; Kurokura et al., 2013).

Recently, studies in *F. vesca* found the homologue of *TFL1* (*FvTFL1*) co-localized with the SFL locus. Unlike the annual model plant *Arabidopsis*, *FvTFL* is in the photoperiod pathway and is seasonally regulated. In DN genotypes, the 2-bp deletion in the coding region of *FvTFL1* caused non-functional *FvTFL1* protein, which lead to activation of continuous flowering and the floral promoter *FvFT1* activates flowering under LD. In the SD genotypes, the decrease of *FvTFL1* expression in autumn, SD photoperiod activated flowering independent of the *FvFT1* pathway. During spring and summer the LD photoperiod promoted *FvTFL1* expression and repressed flowering to continue vegetative growth (Koskela et al., 2012). *FvSOC1* is the *F. vesca* homologue of *Arabidopsis SOC1* and studies found the expression level of *FvSOC1* in seasonal cycling is similar to *FvTFL1* in SD strawberry (Mouhu et al., 2013). Further experiments confirmed *FvSOC1* upregulates *FvTFL1* to repress

flowering under LD. The silencing of *FvSOC1* also caused continuous flowering in LD strawberry (Mouhu et al., 2013).

A previous study based on EST sequencing detected 66 homologues of *Arabidopsis* flowering time genes in diploid strawberry (Mouhu et al., 2009). In that study, gene expression analysis using real-time RT PCR in the DN strawberry 'Baron Solemacher' and 'Hawaii 4' found the *FvAP1* and *FvLFY* expression started to increase after the two-leaf stage in the shoot apical meristem. Genome-scale gene expression profiles of floral development was comprehensively studied using next generation sequencing in *F.vesca* (Hollender et al., 2014). Tissue and stage specific transcriptome profiling in reproductive developmental process were generated as well as modules of coexpressed genes such as meristem regulatory genes, including *LOST MERISTEM* (*FveLOM*) and *WUSCHEL* (*FveWUS*) were identified.

2.4.3 Epigenetic regulation of flowering time

Besides the genetic and environmental factors determining flowering time, DNA methylation influencing flowering induction was studied over the past decades in *Arabidopsis* mutants (Burn et al., 1993; Finnegan, 1996; Russo et al., 1996; Koornneef et al., 1998). These studies indicated hypomethylation of repetitive sequences may inactivate the repressor in the flowering time pathway resulting in early flowering, or reduced DNA methylation in gene regions of FWA caused increased transcription level of FWA and induced late flowering time (Soppe et al., 2000). Previous DNA methylation studies in *Arabidopsis* indicated moderately transcribed genes tended to be methylated rather than the extremely transcribed or non-transcribed genes, and loss of methylation in the gene body promoted gene transcription (Zilberman et al., 2007).

In the vernalization pathway, epigenetic changes are involved in the regulation of flowering time. When plants were exposed to a low temperature environment, the *FLC* gene and other two flanking genes responded to the environmental changes together (Finnegan et al., 2004). Changes in the chromatin at the *FLC* locus also occur such as deacetylation at H3K9 and H3K14 sites, hypermethylation at H3K9 and K27 sites and reduction of trimethylation at H3K4 site (Grant-Downton and Dickinson, 2006). These epigenetic marks led to the silencing of *FLC*, resulting in promotion of flowering, and this memory was stably mitotically transmitted.

Another epigenetic mediated *FLC* repression mechanism is associated with the Methyl-CpG-binding domain (MBD) proteins. The MBD proteins have conserved DNA binding surface targets at symmetrically methylated CpG in DNA regions to regulate the chromatin structure and gene expression (Wade, 2001). In the *Arabidopsis* genome six (AtMBD1, 2, 4, 5, 6 and 7) of a total of twelve putative MBD genes has been detected to bind to the methylated CpG sequence (Berg et al., 2003). The studies of methyl-CpG binding domain (MBD) in *Arabidopsis* found that in AtMBD9 mutants, the reduction of H3 and H4 acetylation level of *FLC* chromatin and global reduction of DNA methylation caused the downregulation of *FLC* expression. This led to early flowering as well as increased axillary branching (Peng et al., 2006; Yaish et al., 2009).

3.0 NOVEL QUANTITATIVE TRAIT VARIATION IS REVEALED IN A HYPOMETHYLATED POPULATION OF WOODLAND STRAWBERRY

3.1 Abstract

Phenotypic variation is determined by a combination of genotype, environment and their interactions. The realization that allelic diversity can be both genetic and epigenetic allows the environmental component to be further separated. Partitioning the variation observed among inbred lines with an altered epigenome can allow the epigenetic component controlling quantitative traits to be estimated. To assess the contribution of epialleles on phenotypic variation and determine the fidelity with which epialleles are inherited, we have developed a novel hypomethylated population of strawberry (2n = 2x = 14) plants using the epimutagen 5-azacytidine from which individuals with altered phenotypes can be identified, selected and characterized.

The hypomethylated population was generated using an inbred strawberry population in the *F. vesca* ssp. *vesca* accession Hawaii 4. To ensure that the observed phenotypic differences did not result from allelic variation, the level of genetic polymorphism in the progenitor and the hypomethylated population was assessed using molecular markers as well as whole genome DNA sequencing. The populations contained only Hawaii 4 alleles, removing introgression of alternate *F. vesca* alleles as a source of variation. Genome sequencing demonstrated that 5-azacytidine did not act as a mutagen where the number of SNP detected among the treated individuals did not increase above spontaneous background levels. Subsequent detailed phenotypic assessments measured quantitative trait variation that focused on flowering time and rosette diameter. The trait mean and variance were determined in both control and hypomethylated populations where we observed expanded levels of variation among the epimutagenized lines. Methylation sensitive molecular markers confirmed that 5-azacytidine induced alterations in DNA methylation patterns and inheritance of methylation patterns were confirmed by bisulfite sequencing of targeted regions. It is possible that methylation polymorphisms might underlie or have induced genetic changes underlying the observable differences in quantitative phenotypes. This population developed in a uniform genetic background of Hawaii 4 provides a resource for the rapid discovery and identification of epialleles controlling quantitative traits. Using this new hypomethylated resource, we have identified new quantitative variation and have demonstrated the inheritance of these phenotypes and variant epialleles through meiosis.

3.2 Introduction

Phenotypic variation is determined by a combination of genotype, environment and their interactions (Michael and Walsh, 1995). Through the use of carefully designed crossing strategies, pedigree analysis and molecular markers, the proportion of the phenotypic variation controlled by genetic components can be estimated and mapped to specific loci with the remaining variation attributed to environmental components (Lamkey and Lee, 1993; Crossa et al., 2010; Melo et al., 2014). Measuring the phenotypic variation observed among genetically uniform individuals allows the extent to which the environmental components affect quantitative traits to be estimated. Populations of inbred lines, the F_1 generation derived from crossing inbred lines and the progeny from double-haploid individuals contain genetically identical individuals and are often used to estimate environmental variance (Johnson et al., 1955; Forster and Thomas, 2005; Hauben et al., 2009; Slovin et al., 2009). However, it is becoming increasingly accepted that epigenetic polymorphism affects quantitative trait variation (Johannes et al., 2009; Reinders et al., 2009; Hu et al., 2015). Epigenetic variation results from the same sequence allele possessing a different chromatin organization, modifying its potential for expression or chromosomal interactions (Feil and Fraga, 2011; Fujimoto et al., 2012). Epigenetic variation has the potential to alter the magnitude of gene expression whereas genetic polymorphisms result from differences in the primary DNA sequence resulting in functional allelic variants. Establishing the effect that epigenetic variation has on quantitative characters is complicated by any underlying genetic variation (Richards, 2011) and it is often considered a component of environmental variation (Ramchandani et al., 1999; Angers et al., 2010; Feil and Fraga, 2011). However, the development of tools that are able to detect and measure epigenetic variation in genetically uniform populations is allowing these issues to be resolved (Johannes et al., 2009; Verhoeven, 2010).

Epigenetic information is stored in three molecular systems, namely, DNA methylation, post-translational modification of histone proteins and non-coding RNA molecules. Together, these systems organize the structure and configuration of chromatin adjusting its accessibility to the transcriptional machinery that can result in the activation or repression of gene expression (Saze, 2008; Feng et al., 2010). Perhaps the best studied of these systems is DNA methylation (Bird, 2002; Reinders et al., 2009). In eukaryotes, DNA can be modified through the covalent attachment of a methyl group to the carbon atom at position five of the cytosine

ring, a reaction catalyzed by methyltransferases (Bestor et al., 1988; Bird, 1992). In contrast to animals, where cytosine methylation is largely restricted to the CG sequence context, DNA methylation in plants is additionally observed at positions with CHG and CHH sequence contexts (where H = A, C, T) (Oakeley and Jost, 1996). DNA methylation status is maintained through DNA replication through the action of DNA maintenance enzymes such as DNA METHYLTRANSFERASE 1 (MET1) and CHROMOMETHYLASE 3 (CMT3) transferring the methylation status to the newly synthesized stand using symmetry in the CG and CHG sequence contexts (Lindroth et al., 2001; Kankel et al., 2003). Whereas the action of *de novo* methylation enzymes including DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) are required to maintain the information at asymmetric sites CHH where siRNA molecules act as guides providing the necessary sequence specificity (Goll and Bestor, 2005; Law and Jacobsen, 2010). Major differences in cytosine methylation patterns throughout plant genomes distinguish gene-rich regions from repeat-rich regions but variation in methylation among genes and their promoters has been shown to control expression (Bender, 2004). Altering the constraints provided by DNA methylation to gene expression might manifest as changes in phenotypes and offers a mechanism for epigenetic control over quantitative traits. Additionally, alterations in DNA methylation patterns have the potential to be indirectly mutagenic, affecting the higher-order chromosome conformation leading to chromosomal rearrangements that might also lead to trait variation (Wolffe and Matzke, 1999; Martienssen and Colot, 2001; Bell et al., 2011; Bell et al., 2012).

Although comparatively rare, examples of naturally occurring heritable traits controlled by epialleles are known. This is best exemplified with the demonstration that changes in DNA methylation patterns are responsible for the non-Mendelian inheritance of floral symmetry observed in Toadflax. These epigenetic alterations cause silencing of *Lcyc* gene expression, resulting in the easily observable change from bilateral to radial floral symmetry (Cubas et al., 1999). Additionally, fruit ripening in tomato was inhibited by a spontaneous epimutation resulting from hypermethylation of the SBP-box transcription factor promoter sequence (Manning et al., 2006). Transgenerational fidelity of epigenetic information in plants has been demonstrated more generally and there is a growing body of evidence describing the importance of these phenomena (Kakutani et al., 1999; Heard and Martienssen, 2014). Stable inheritance of naturally occurring epialleles was demonstrated through genome-wide analyses in studies using *Arabidopsis* where the majority of DNA methylation patterns were faithfully

transmitted through meiosis for many generations, dispelling the idea that these marks were largely transitory and subject to environmental change (Schmitz et al., 2011). Stable inheritance of phenotypic variation in flowering time and plant height induced by altering DNA methylation patterns was observed for over eight generations in *Arabidopsis* using a population of epigenetic Recombinant Inbred Lines (epiRILs) suggesting that quantitative trait variation might also be under epigenetic control (Johannes et al., 2009). The variation among the epiRILs was generated in isogenic backgrounds through the functional inactivation of the maintenance methyltransferase resulting in lines with reduced DNA methylation primarily at CG sites (Johannes et al., 2009).

Established patterns of DNA methylation can be stochastically changed through the application of potent pharmacological agents such as 5-azacytidine (5-azaC) (Veselý and Čihák, 1978; Veselý, 1985). This compound is an analogue of cytidine possessing a nitrogen atom rather than a carbon atom at the 5-position of the pyrimidine ring and thus lacking the ability to form a bond with a methyl group. During DNA replication, 5-azaC competes with cytosine as a substrate for DNA polymerase and is incorporated into the genome resulting in demethylation as marks cannot be transferred by maintenance methylase enzymes to the newly synthesized DNA strand (Christman, 2002; Stresemann, 2008). Studies altering genome-wide DNA methylation patterns have demonstrated that exposure to exogenous 5-azaC in plants can increase phenotypic trait variation. Strategies exploiting this phenomenon have been applied in a range of plant species where it has been demonstrated that dwarfism in *Oryza sativa* (Sano et al., 1990; Akimoto et al., 2007), early flowering in *Arabidopsis* (Burn et al., 1993), *Linum usitatissimum* (Fieldes, 1994b; Fieldes and Amyot, 1999; Fieldes et al., 2005), and *Solanum ruiz-lealii* (Marfil et al., 2012), as well as abnormal flower development, and leaf morphology *Solanum ruiz-lealii* (Marfil et al., 2009; Marfil et al., 2012) have been identified.

The Rosaceae family contains the majority of the most economically important temperate fruit crops such as apple, cherry, pear, plum, peach, raspberry, almond as well as strawberry (Hummer and Janick, 2009). Woodland strawberry (*F. vesca*, 2n=2x=14) has a rapid generation time, small plant stature, is able to reproduce through both sexual and clonal pathways and has a relatively small genome (~240Mb). The major advantage that *F. vesca* offers for functional genomics is the availability of the genome sequence generated using the inbred line Hawaii 4 of the *F. vesca* ssp. *vesca* (Shulaev et al., 2011). Together, these features

have transformed this species into the model species for both strawberry and the wider Rosaceae family. The generation of the genome sequence paves the way for genomics analyses to determine the function of strawberry genes and is a prerequisite for detailed investigations describing epigenetic variation.

Here we describe the generation of a new resource that can be used in future analyses to address fundamental questions of epigenetic gene regulation and its contribution to quantitative phenotypic variation in *F. vesca.* This novel resource was developed using the Hawaii 4 genetic background with the prospective of generating and discovering novel factors (alleles or epialleles) that affect trait variation. We assess the extent of observed changes in DNA methylation patterns and phenotypic variation in these plants and assess whether this variation is transmitted through meiosis.

3.3 Materials and Methods

3.3.1 Plant materials

The *F. vesca* ssp. *vesca* accession Hawaii 4 (germplasm accession: PI551572) was used in this study and seeds obtained after seven generations of inbreeding (H4S7) derived through single-seed descent were kindly provided by Dr. Janet Slovin (USDA). These seeds were direct descendants of the individual plant used to generate the *F. vesca* reference genome (H4S4). The plant material used for population development was derived from seeds generated from an additional generation of inbreeding (H4S8) required to produce sufficient material for mutagenesis.

3.3.2 Exposure of *F. vesca* to 5-azacytidine

A population of H4S8 seeds was treated with 5-azaC (Sigma-Aldrich, St. Louis, MO, USA). Seeds were imbibed with water for 24 hours and then treated with 0, 1.0, 5.0, 20.0, 50.0 or 100.0 mM of 5-azaC and incubated at room temperature in the dark for six days before being rinsed and germinated on filter paper in petri dishes. Germinated seedlings were transferred to pots containing Sunshine Mix #4 (Sun Gro Horticulture), and placed on benches under greenhouse conditions where they were grown at 23 ± 2°C day and 18 ± 2°C night under an 18 / 6 h day / night photoperiod. Natural light was supplemented with 400W high-pressure sodium lights at 600 μmol m^{-2} s^{-1}. Plants were fertilized once per week with 2 g / L of NPK (20-20-20) including micronutrients (Plant Products Co. Ltd., Brampton, ON, LGT-1J1).

3.3.3 Phenotypic traits assessments

The 5-azaC treated and control plants were assigned a code as a unique accession identifier (ERFv#) to ensure phenotypic assessment was conducted without bias. The 5-azaC treated and control populations were scored for two phenotypic traits: (1) Flowering time, recorded as the number of days from sowing to anthesis, determined by the opening of the first (primary) flower; (2) Rosette diameter (mm), recorded as the maximal linear distance across the strawberry rosette 45 days after sowing.

3.3.4 Preparation of strawberry genomic DNA

Fresh leaf material was harvested directly into liquid nitrogen and stored at -80°C until DNA extraction. DNA from ~400 mg of leaf material was extracted using the CTAB (cetyltrimethyl ammonium bromide) method (Murray and Thompson, 1980) with the following modifications. To obtain high-quality DNA from strawberry, the tissue was ground in liquid nitrogen to a fine powder and transferred to a sorbitol buffer (100 mM pH 8.0 Tris-HCl, 0.35 M sorbitol, 5 mM pH 8.0 EDTA, 1% PVP-40 with 1% 2-mercaptoethanol) which was used as a wash buffer to remove excessive mucilaginous polysaccharides prior to CTAB extraction (Souza et al., 2012). DNA quantification was performed using Qubit 2.0 Fluorometer and the Qubit dsDNA BR Assay Kits (Invitrogen, Eugene, OR, USA) according to the manufactures instructions.

3.3.5 Assessment of genetic variation using Amplified Fragment Length Polymorphism (AFLP)

The AFLP protocol described by Vos et al., (1995) was followed with modifications: A total of 250 ng of genomic DNA extracted from leaf material was digested with 10 units *Eco*R I, 5 units *Mse* I (New England Biolabs), in NEB-4 buffer with BSA in a final volume of 40 µl for 2 hours at 37 °C and the enzymes denatured by incubation for 15 minutes at 70°C. Ligation of adaptors (**Table 3.1**) to the *Eco*R I and *Mse* I digested DNA was performed using NEB-4 buffer, BSA, ATP and 100 cohesive end units of T4 DNA ligase (New England Biolabs) in a total volume of 50 µl at room temperature for 2 hours. The ligation reaction was diluted 1:10 before AFLP pre-selective amplification. Pre-selective PCR reactions was performed in a volume of 50 µl containing 5 µl of 1:10 ligation dilution, 0.1µM of the *Eco*R I and *Mse* I primers (**Table 3.1**), 1× PCR buffer with $MgCL_2$, 200 µM dNTP and 1 unit of Taq polymerase. The conditions for pre-selective PCR were as follows: 19 cycles of 94°C for 30 s,

56°C for 1 min, and 72°C for 1 min.

The product of pre-selective amplification was diluted 1:50 and used as template in the selective amplification reaction. Selective amplification reactions were performed in a final volume of 25 µl containing 5 µl of the diluted pre-selective amplification product, 0.05 µM ^{33}P labeled *Eco*R I selective primer (**Table 3.1**), 0.25 µM *Mse* I selective primer (**Table 3.1**), 1× PCR buffer, 200 µM dNTP and 1 unit of Taq polymerase. The conditions for selective PCR were as follows: 12 cycles of 94°C for 30 s, 65°C for 30s, and 72°C for 1 min, then followed by 22 cycles of 94°C for 30 s, 56°C for 30s, and 72°C for 1 min.

3.3.6 DNA sequencing libraries construction

Whole genome DNA sequencing was conducted in three control lines (Con1, Con2, Con3) and four 5-azaC treated lines (ERFv140, ERFv246, ERFv132, ERFv153). Illumina TruSeq DNA libraries were prepared following the manufactures' instructions. Briefly, one µg of whole genomic DNA was sheared using the Bioruptor (Diagenode) using 12 cycles, pulsing for 30 seconds with 190 seconds gap between pulses. Following fragmentation, end repair, and adapter ligation, the BluePippin Prep (Sage Science) was used to capture 590 bp fragments. The libraries were quantified using the 2100 Bioanalyzer (Agilent Technologies) and sequencing was performed using Illumina HiScanSQ platform according to the manufacturer's instructions.

3.3.7 Sequence Alignment, Single Nucleotide Polymorphism (SNP) identification, and SNP annotation

Sequence reads in fastq format were filtered and trimmed using Trimmomatic v0.32 (Bolger et al., 2014). Sequence quality assessment was conducted using CLC Genomic Workbench 8.5. The filtered libraries were aligned to the reference *F. vesca* whole genome (v1.0) (Shulaev et al., 2011) using CLC Genomic Workbench 8.5. Variant calling was performed using HaplotypeCaller and SNP identification was performed using SelectVariants with Genome Analysis Toolkit (GATK) (McKenna et al., 2010). Further filtering of SNP variant calls was performed using custom Perl scripts where high confidence SNP were identified by selecting for those loci with at least three reads, that were not adjacent to an identified Indel (adjacency was determined by the length of the detected Indel) and did not share a common genotype in control and 5-azaC treated samples. SNP annotation and functional prediction of

Table 3.1 Sequences of adaptors and primers used for pre-selective amplification and selective amplification in AFLP.

Adapters / Primers	*Eco*R I	*Mse* I
Adapter	5-CTCGTAGACTGCGTACC-3	5-GACGATGAGTCCTGAG-3
	3-CATCTGACGCATGGTTAA-5	3-TACTCAGGACTCAT-5
Pre-amplification primers	5-GACTGCGTACCAATTC A-3	5-GATGAGTCCTGAGTAA C-3
Selective primers	5-GACTGCGTACCAATTC AC-3	5-GATGAGTCCTGAGTAA CAA-3
	5-GACTGCGTACCAATTC AC-3	5-GATGAGTCCTGAGTAA CAT-3
	5-GACTGCGTACCAATTC AC-3	5-GATGAGTCCTGAGTAA CCA-3
	5-GACTGCGTACCAATTC AC-3	5-GATGAGTCCTGAGTAA CGT-3

the variants were performed using SnpEff (Cingolani et al., 2012) based on the annotations provided by the *F.vesca* genome v1.0 (Shulaev et al., 2011). The protein sequences for the set of *Arabidopsis* flowering time genes listed at http://www.mpipz.mpg.de/14637/Arabidopsis_flowering_genes (Fornara et al., 2010) obtained from TAIR were used to identify putative *F. vesca* flowering time homologues through sequence alignment.

3.3.8 Assessment of DNA methylation polymorphism using Methylation Sensitive Amplified Polymorphisms (MSAP)

The MSAP protocol was followed with slight modification of the original protocol (Reyna-López et al., 1997). Briefly, genomic DNA from each sample analyzed was digested separately with 10 units *Eco*R I / 5 units *Hpa* II (New England Biolabs) and 10 units *Eco*R I / 10units *Msp* I (New England Biolabs). The *Eco*R I and *Hpa* II / *Msp* I adaptors (**Table 3.2**) were annealed and ligated to digested DNA fragments. The pre-selective and selective primers were listed in **Table 3.2**. Amplification of DNA fragments for MSAP followed the same PCR cycling conditions used for AFLP.

3.3.9 Resolution and Scoring of amplified AFLP and MSAP products

The selective PCR amplification products from AFLP and MSAP were resolved using a 5% polyacrylamide gel using the BioRAD Sequi-Gen vertical polyacrylamide gel system. The resulting gel was dried and exposed to autoradiographic film (Kodak BioMax MR film 35 X 43 cm). The size of the visible fragments was determined using ^{33}P labeled 50 bp ladder. A total of four and ten primer pairs were used to assay for polymorphism using the AFLP and MSAP method respectively (**Table 3.1, Table 3.2**).

Scoring of the AFLP and MSAP data were restricted to the clearly amplified fragments and data were recorded as dominant allelic markers. The banding patterns representing each observed allele in each individual were encoded by single and double band values for the AFLP and MSAP data respectively. In **Figure 3.1** for each locus, if there were bands resolved after electrophoresis in both *Eco*R I / *Hpa* II and *Eco*R I / *Msp* I digest lanes, it was scored as 1 / 1 (type I band). In this situation, cytosine was not methylated. When the bands were present in the *Eco*R I / *Hpa* II digest and absent in the *Eco*R I / *Msp* I digest it indicated that cytosine methylation was present on one strand of the DNA, called hemimethylation and

scored as 1 / 0 (type II band). Although hemimethylation can occur in both external and internal cytosines or only in the external cytosine of the 5'-CCGG-3' recognition sequence, the former pattern is of higher frequency. When the bands were present in the *Eco*R I / *Msp* I digest and absent in *Eco*R I / *Hpa* II digest, the internal cytosines in both strands were methylated and scored as 0 / 1 (type III band). If there were no bands, it was scored as 0 / 0 (type IV band) showing fully methylation, and both internal and external cytosine methylation patterns having a higher frequency compared to only external cytosine methylation form.

3.3.10 Generation of high-resolution DNA methylation patterns at target loci

Putative CG islands in the *F. vesca* genome were identified using a custom Perl Script. CG islands were defined as being a minimum of 300 bp with greater than 50% GC content and an observed-CG / expected-CG ratio greater than 0.6. Three target regions were selected using the *F. vesca* genome v1.0 to determine the methylation patterns that are enriched for cytosine bases. Target region one is on chromosome one between positions 1107633 and 11077319; Target region two is on chromosome two between positions 1029956 and 1030513 and Target region three is on chromosome four between positions 4884809 end 4885267. To ensure efficient amplification PCR primers were designed to amplify products less than 500 bp since conversion using sodium bisulfite can degrade the integrity of genomic DNA (Henderson et al., 2010). To ensure efficient annealing, primers were designed to avoid the presence of cytosine bases making them able to amplify from sequences possessing either methylated or unmethylated cytosine bases. The primer sequences used are presented in **Table 3.3**.

Genomic DNA was treated with sodium bisulfite using the EZ DNA Methylation-Gold Kit (Zymo Research), by incubation at 98°C for 10 min, 64 °C for 2.5 h in a thermal cycler. Lambda DNA (150 ng) was spiked into each sample as an unmethylated reference to calculate conversion rate efficiency. The converted DNA was used as template DNA in the PCR to amplify target genomic loci. The PCR was performed in 50 µl final volume with Zymo*Taq*™ Premix 25 µl (Zymo Research), 5 µl of each primer (10 µM), template DNA and H_2O 15 µl. The conditions for PCR were as follows: 95 °C for 10 min followed by 40 cycles of 95 °C for 30s, 55°C for 40s, and 72°C for 60s with a final extension step at 72 °C for 7 min. The PCR product was sequenced and aligned to the reference sequence using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). CyMATE was used for visualization to detect the methylation patterns (http://cymate.org/cymate.html) (Hetzl et al., 2007).

Table 3.2 Sequences of adaptors and primers used for pre-selective amplification and selective amplification in MSAP.

Adapters / primers	*Eco*R I	*Hpa* II / *Msp* I
Adapter	5-CTCGTAGACTGCGTACC-3	5-GACGATGAGTCTAGAA-3
	3-CATCTGACGCATGGTTAA-5	3-CTACTCAGATCTTGC-5
Pre-amplification primers	5-GACTGCGTACCAATTC A-3	5-GATGAGTCTAGAACGG T-3
Selective primers	5-GACTGCGTACCAATTC ACT-3	5-GATGAGTCTAGAACGG TAA-3
	5-GACTGCGTACCAATTC ACT-3	5-GATGAGTCTAGAACGG TTA-3
	5-GACTGCGTACCAATTC ACT-3	5-GATGAGTCTAGAACGG TCA-3
	5-GACTGCGTACCAATTC ACT-3	5-GATGAGTCTAGAACGG TGA-3
	5-GACTGCGTACCAATTC ACT-3	5-GATGAGTCTAGAACGG TTC-3
	5-GACTGCGTACCAATTC ACT-3	5-GATGAGTCTAGAACGG TTG-3
	5-GACTGCGTACCAATTC ACT-3	5-GATGAGTCTAGAACGG TTT-3
	5-GACTGCGTACCAATTC ACA-3	5-GATGAGTCTAGAACGG TAA-3
	5-GACTGCGTACCAATTC ACA-3	5-GATGAGTCTAGAACGG TTA-3
	5-GACTGCGTACCAATTC ACA-3	5-GATGAGTCTAGAACGG TTT-3

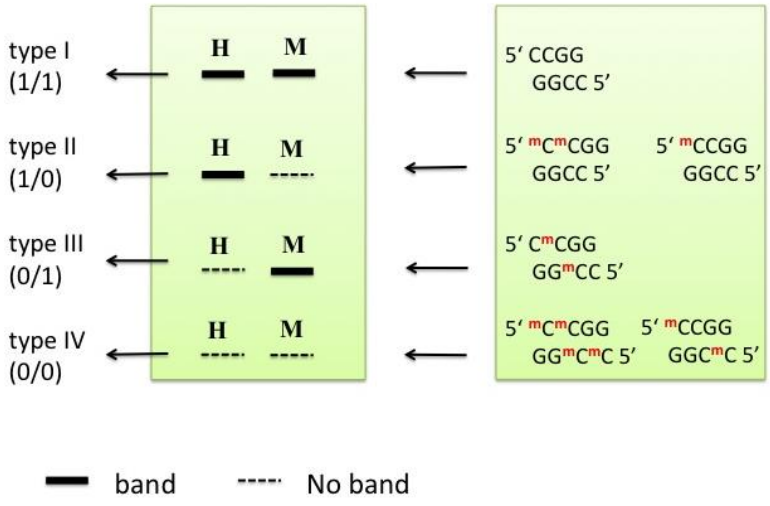


Figure 3.1 *Hpa* II (H), *Msp* I (M) sensitivity to methylation at 5'-CCGG-3' sites and scoring of MSAP bands (modified from Salmon et al., 2008).
The type IV band represents full methylation, the fully methylated 5'mCCGG sequence was added here. The "1" represents the presence of bands and "0" represents the absence of bands for scoring purposes.

Table 3.3 Primers used in bisulfite sequencing PCR.

	Forward Primer Sequence	Reverse Primer Sequence
Target region one	5-GAGTAGAGGGTGGTTGATT-3	5-TCTCACCAACCAATCAAATC-3
Target region two	5-GAGGATGATATTGTGAGGTTA-3	5-CTACACCCTATAAACACCTA-3
Target region three	5-TTGAGTAGAAGTGAATGAG-3	5-AACCTCCAAACTACTCTCTA-3
Lambda primer	5-GATGTGTAGGTTATGGTGA-3	5-CCATCACACTCACATAAC-3

3.3.11 Statistical analysis

The quantitative phenotypic data obtained from measuring flowering time and rosette diameter were analyzed using the statistical software R. Basic descriptive statistics including the mean and variance were estimated from the control population for each character, and the significance of each deviation from the control population mean was determined using one sample Z-test.

MSAP profiles describing the methylation patterns observed in the 5-azaC treated and control lines were summarized by Principal Coordinates Analysis (PCoA) and compared by Analysis of Molecular Variance test (AMOVA) using the MSAP analysis package for R (Pérez-Figueroa, 2013). Loci with at least 5% methylated levels were defined as methylation-susceptible loci. Polymorphic methylation-susceptible loci were defined when at least two individuals were non-methylated (Herrera and Bazaga, 2010; Morán and Pérez-Figueroa, 2011).

3.4 Results

3.4.1 Generation of a hypomethylated population of *F. vesca*

Approximately 500 *F. vesca* seeds from generation H4S8 were exposed to a range of 5-azaC concentrations (0-100 mM). A total of 305 plants survived 5-azaC treatment and transplantation into soil. These were complemented with a population of 59 H4S8 control plants. The surviving treated population was composed of plants exposed to a range of 5-azaC concentrations (1.0, 5.0, 20.0 or 50.0 mM), whereas the control population was exposed to water. Seeds exposed to 5-azaC concentrations above 50 mM were unable to survive and the largest class, comprising ~40% of the population, was exposed to 20 mM 5-azaC.

3.4.2 Genetic uniformity was verified in epimutagenized population

Since any genetic polymorphism is likely to complicate the effects resulting from induced epigenetic differences (Eichten et al., 2013), in order to attribute any phenotypic variation observed in quantitative characters to epigenetic variation, it is necessary to perform the experiment using a genetically uniform population (Richards, 2011). The Hawaii 4 lines used in this study were highly inbred, derived through single-seed descent for a total of eight generations (H4S8). This level of inbreeding strongly suggested the material used to develop

the epimutant population was genetically uniform. This was initially confirmed using AFLP markers to assess the genetic background *F. vesca* material used in this study. Genotyping was performed using a randomly selected subpopulation consisting of five control lines and 22 lines from the epimutagenized population. Although AFLP markers are dominant, the large number of loci amplified per primer pair means they can be used to quickly survey the entire genome (Mueller and Wolfenbarger, 1999). A total of 219 AFLP loci were amplified using four primer pair combinations (**Table 3.1)**. The allelic banding patterns observed were identical throughout all of the 27 individuals examined, indicating that no introgression of alien *F. vesca* alleles had inadvertently occurred through hybridization during inbreeding (**Figure 3.2**).

Further evidence of genetic uniformity was achieved by whole genome sequencing of selected lines to address the potential of 5-azaC to act as a mutagen. A total of seven lines, comprising three untreated and four 5-azaC treated lines were sequenced which resulted in the generation of 29,569,617 sequence reads. After exclusion of sequences comprising low quality bases a total of 21,502,412 sequence reads were aligned to the reference genome of *F. vesca* (H4S4). The sequence alignment resulted in 187 MB (~90%) coverage of the genomes with the depth ranging from 8 to 14 (**Table 3.4, Table 3.5**). SNP loci were identifed from the short read sequence alignments for each of the seven lines using GATK. High confidence SNP detection required evidence for an alternate allele from at least three independent sequence reads, with a SNP quality score of >3000. Additionally, SNP loci were excluded when positioned adjacent to an identified Indel as these are likely to result from alignment artefacts.

The total number of loci with high quality SNP was 30685 where an alternate allele was present in at least one of the seven lines. However, for the vast majority of these loci (29137 (95%)), although the H4S8 allele differed from the reference allele, the genotype of all seven of the H4S8 lines was identical. The remaining 1548 (5%) loci where variation among the seven lines was observed were further partitioned. A total of 1208 (4%) loci were heterozygous and 340 (1%) of the total loci possessed homozygous alternative alleles in at least one of 5-azaC treated lines and this genotype was absent from the control lines. Further inspection of the data revealed 153 (0.5%) loci where greater than one alternative allele was detected among the seven lines. Among these, both alternate alleles were found in the control and 5-azaC treated lines for 148 (97%), these might result from spontaneous mutation. At the remaining 5 (3%) loci the alternate alleles where found exclusively among the 5-azaC treated

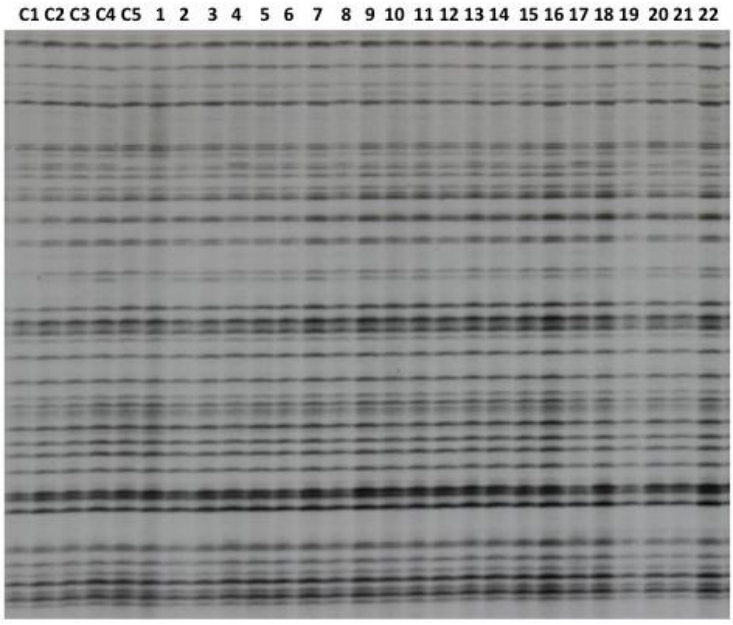


Figure 3.2 Example of AFLP electrophoretic patterns in control and epimutant lines. Banding pattern was generated using the primer combination *Eco*R I + AC and *Mse* I + CAA. C1-C5: five control lines. 1-22: randomly selected 22 hypomethylated lines.

Table 3.4 The distribution of coverage in each line across the genome.

	Reference	Con 1	Con 2	Con 3	ERFv 140	ERFv 246	ERFv 132	ERFv 153
LG1	24,178,039	21,317,245	21,350,972	21,317,109	21,331,069	21,361,619	21,366,204	21,348,926
LG2	35,504,443	31,197,786	31,248,980	31,199,229	31,222,325	31,265,870	31,275,762	31,246,727
LG3	29,376,571	26,342,781	26,385,557	26,345,840	26,362,986	26,400,600	26,407,963	26,383,537
LG4	24,141,177	21,802,964	21,840,725	21,805,867	21,817,819	21,851,261	21,857,516	21,837,786
LG5	30,775,793	27,706,440	27,756,275	27,712,527	27,728,668	27,769,668	27,777,504	27,749,944
LG6	39,918,795	35,990,843	36,052,431	35,998,091	36,022,429	36,072,557	36,080,868	36,049,501
LG7	24,052,591	22,234,862	22,271,854	22,239,518	22,253,469	22,281,798	22,285,625	22,266,143
All	207,947,409	186,592,921	186,906,794	186,618,181	186,738,765	187,003,373	187,051,442	186,882,564

Table 3.5 The distribution of depth of coverage in each line across the genome.

	Con 1	Con 2	Con 3	ERFv 140	ERFv 246	ERFv 132	ERFv 153
LG1	7.94	10.5	7.8	8.72	11.73	13.69	9.46
LG2	7.7	10.2	7.6	8.47	11.48	13.32	9.34
LG3	8.22	11.03	8.14	9.13	12.24	14.37	9.85
LG4	8.01	10.8	8.03	8.95	12.24	14.09	9.93
LG5	8.94	10.7	7.92	8.89	11.97	14.01	9.66
LG6	8.22	12.06	8.9	9.92	13.45	15.62	10.6
LG7	8.25	10.96	8.1	9.08	12.25	14.28	9.95
Average	8.18	10.89	8.07	9.02	12.19	14.20	9.83

individuals. The slight increase in allelic complexity (3% increase in the loci possesing an extra allele) observed among the 5-azaC treated compared to control lines was not statistaically significant when testing for an increase in the propotion of loci with greater than two alleles ($\chi^2 = 0.0589$; $p = 0.8083$). The biological significance of the detected SNP was assessed by assigning functional annotation to the SNP loci using SnpEff classified using the *F. vesca* genome annotation (Shulaev et al., 2011). The majority 1365 (88%) of the 1548 polymorphic loci identified were found in intergenic regions with only 183 (12%) of the loci annotated as being in genes. Base changes resulting in the predicted loss of gene function accounted for 11 (0.7%) of the loci, 119 (8%) loci predicted nonsynonymous bases changes and 53 (3%) loci resulted in prediction of synonymous base changes. No SNP were detected in *F. vesca* homologues of the flowering time genes in *Arabidopsis*. In the case of those loci with SNP variation found exclusively in the 5-azaC treated lines, six were predicted to result in loss of gene function, 46 nonsynonymous and 14 synonymous substitutions with 346 annotated in intergenic regions.

3.4.3 Expanded Phenotypic variation was observed in the 5-azaC treated population

A number of phenotypic differences were observed upon visual examination of the individuals comprising the 5-azaC population. Quantitative characters including flowering time and plant rosette diameter were scored (**Figure 3.3**). These data were summarized using descriptive statistics and the distributions were visualized for each subpopulation of plants, exposed to different concentrations of 5-azaC (**Figure 3.4**). Individuals were selected that were significantly different from the mean of the control phenotypic values for flowering time and rosette diameter. The population distribution for each phenotype scored was summarized (**Figure 3.4 A**). These distributions indicate the effect of 5-azaC treatment resulted in expanded variation anticipated for these quantitative traits. Treatment with low concentrations of 5-azaC (1.0 mM and 5.0 mM) caused little deviations from those observed among the control lines for flowering time. Treatment with higher levels of the mutagen (20 mM and 50 mM) was required to induce a wide range of phenotypic variation at both tails of the distribution (**Figure 3.4 A**). The median flowering time of 20.0 mM treatment was one day earlier than the control while the average flowering time for 50.0 mM was three days later. As anticipated, the greatest variation was observed among those plants exposed to the treatments with the highest 5-azaC concentration. Similar to the data collected describing flowering time, exposure to 5-azaC treatment increased the variance for rosette diameter. Rosette diameter appeared to be more susceptible to alterations induced by 5-azaC as these lines tended to be

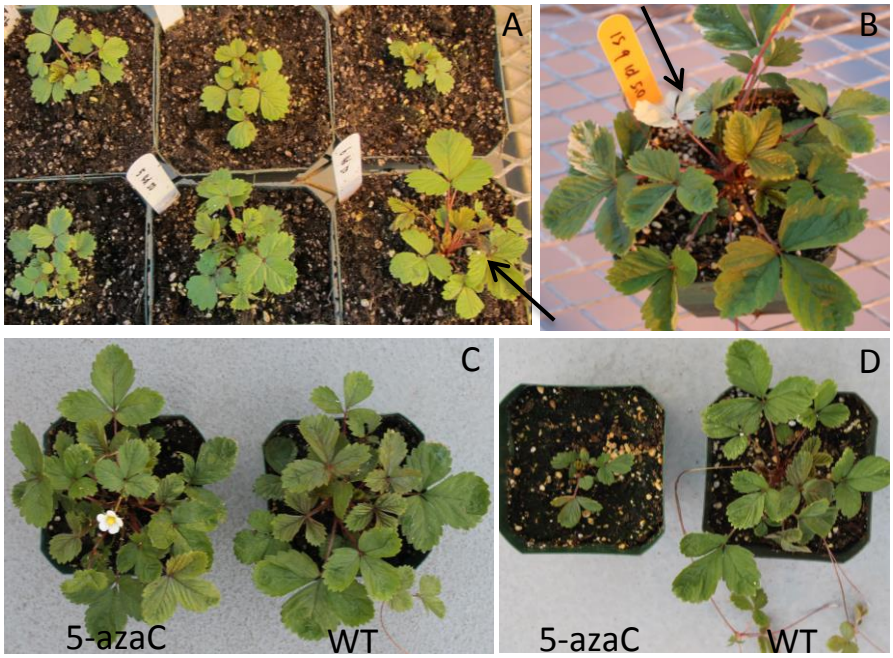
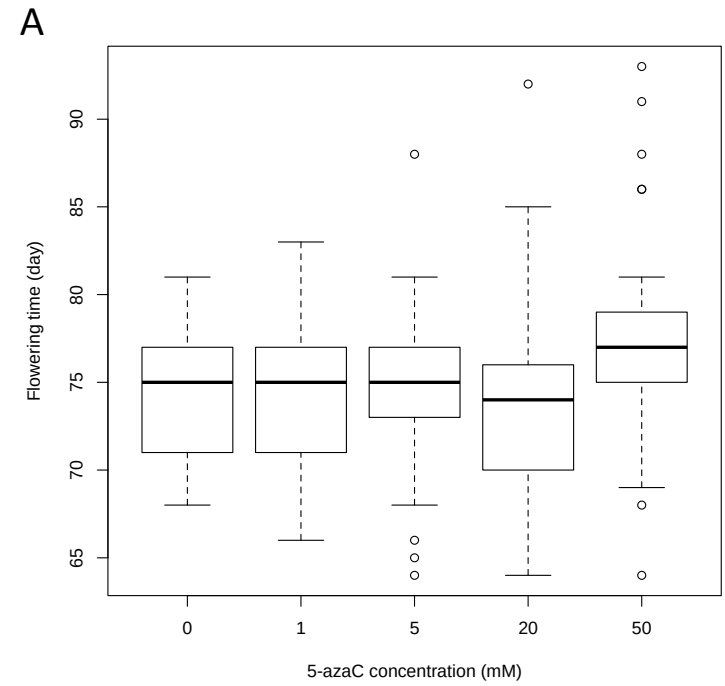


Figure 3.3 Phenotypic variation in plants subjected to 5-azaC treatments under greenhouse conditions.
A: An example of six individual plants varying in plant growth and development parameters under 50mM 5-azaC treatment, right bottom plant ERFv 153 showed early flowering (see arrow). B: Variegated chlorophyll pigmentation was observed in the progeny of one late flowering line ERFv 134 (see arrow). C: The comparison of flowering time between wild type (right) and 5-azaC treatment line (left). D: The comparison of plant rosette diameter and stature between wild type (right) and 5-azaC treatment line (left).



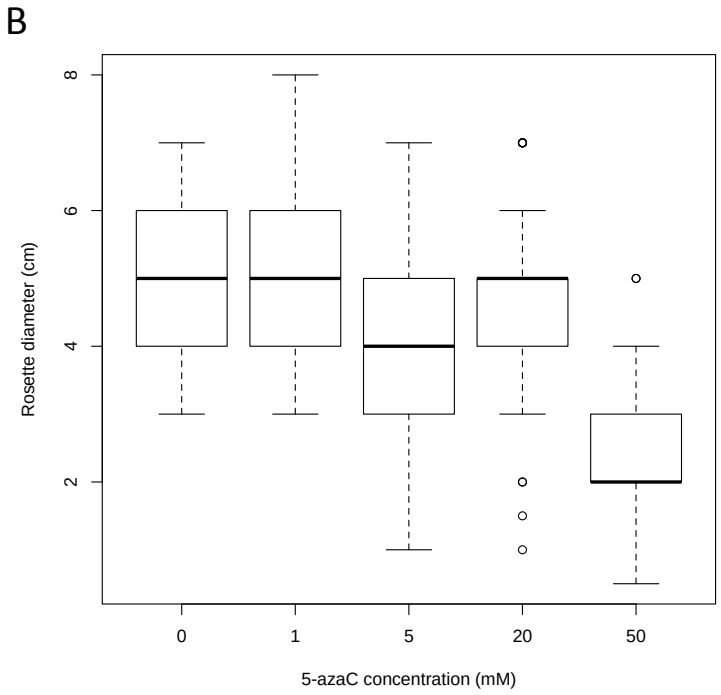


Figure 3.4 Distribution of quantitative phenotypes for control and epimutagenized lines. A: Phenotypic variation observed for flowering time. B: Phenotypic variation observed for Rosette diameter.

smaller than the diameter measured among the control lines (**Figure 3.4 B**). This was particularly evident among those plants exposed to the higher concentrations (20 mM, 50 mM) of 5-azaC. The rosette diameters of those plants exposed to 50 mM 5-azaC were the smallest and the distribution was skewed to smaller diameters, rather than possessing outliers at both tails of the distribution as observed for flowering time.

Ranking of phenotypic values identified those individuals with the most variant phenotype. There were a total of five (1.7%), nine (3.1%), one (0.3%) and 28 (9.8%) individuals identified that presented as having a significantly different ($p < 0.05$) early flowering time, late flowering time, large diameter or small diameter respectively (**Figure 3.5, Table 3.6**). The majority of these variant phenotypes were observed among the lines exposed to higher concentrations of 5-azaC (**Table 3.6**).

3.4.4 Alteration in DNA methylation pattern was detected in the epimutagenized population

Exposure to 5-azaC is known to cause alterations in DNA methylation. To assess the efficacy of the drug we surveyed the same subpopulation used for the AFLP analysis, assaying for variation in cytosine methylation status using the MSAP protocol. A total of 246 MSAP loci were amplified in all individuals using ten primer pair combinations. In contrast to the AFLP profiles observed, the MSAP banding patterns revealed a range of polymorphic loci. The MSAP patterns reveal distinct cytosine methylation status at the sampled 5'-CCGG-3' restriction sites (**Figure 3.6**). These patterns were grouped into four classes allowing estimates of DNA methylation, at sampled sites to be generated (**Table 3.7**). As anticipated, the level of DNA methylation observed was negatively related to the concentration of 5-azaC the plants were exposed (**Table 3.7**). These data indicate 5-azaC reduced DNA methylation within the treated individuals. Although a significant reduction in DNA methylation was observed upon exposure to 5-azaC when compared to controls (**Table 3.8**), the major contribution to this reduction ($p < 0.05$) occurred when the concentration was increased to 50 mM (**Table 3.7**). The MSAP data revealed that the changes induced by 5-azaC are most prevalent in the type I (unmethylated class) and type IV (fully methylated class). Exposure to 5-azaC increases the frequency of loci that possess no methylated cytosine (45% to 52%) and decreases the frequency of 5'-CCGG-3' sites with four methylated cytosine bases (21% to 14%). The frequency of the type II MSAP banding pattern (hemimethyated) showed no change and the type III (internal base methylation) showed a minor reduction in frequency. Examination of the methylated cytosine bases (Type II + Type III + Type IV, **Table 3.7**)

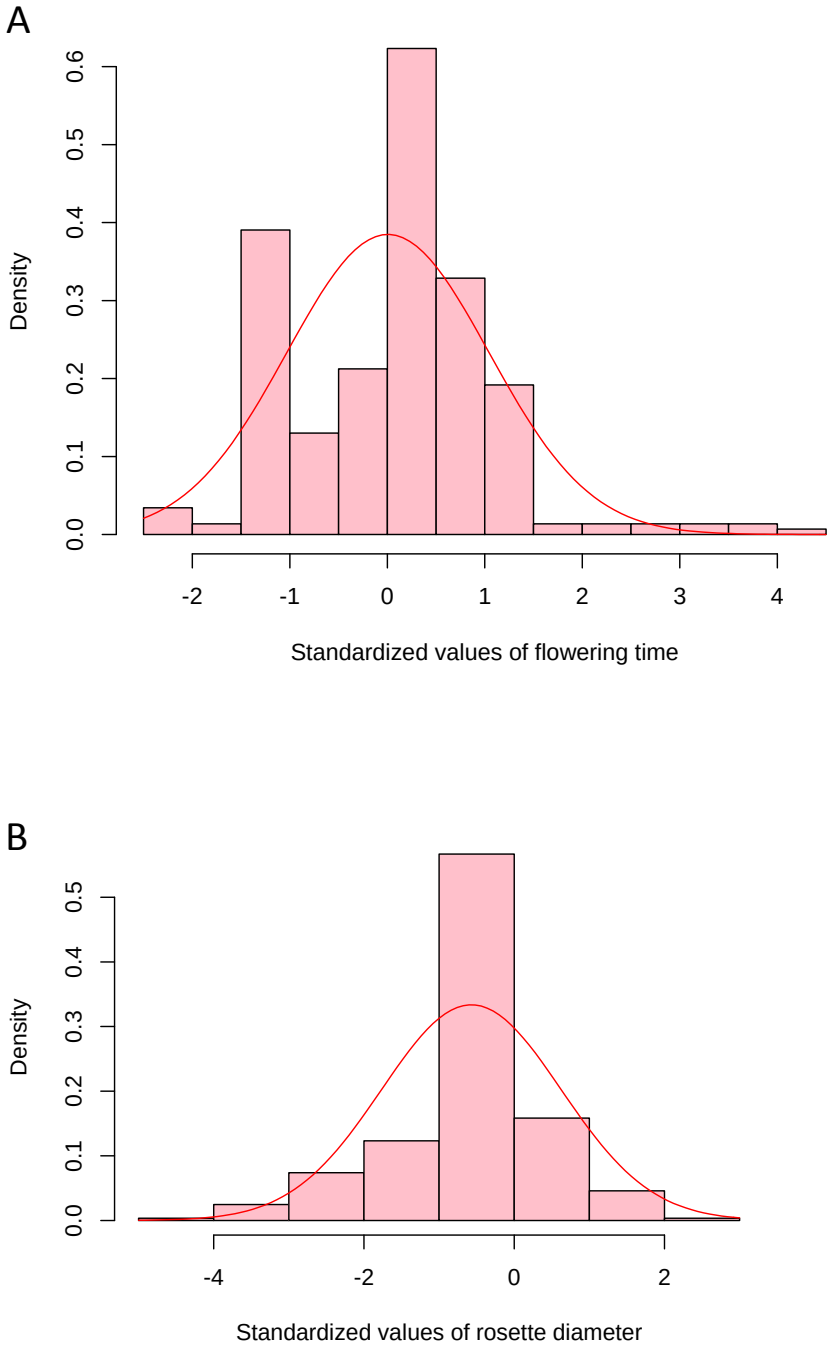


Figure 3.5 Distribution of standardized values for two quantitative traits observed in the epimutangenized population relative to the control population.
A: Flowering time density histogram of Z-test values from 292 hypomethylated lines.
B: Rosette diameter density histogram of Z-test values from 284 hypomethylated lines.

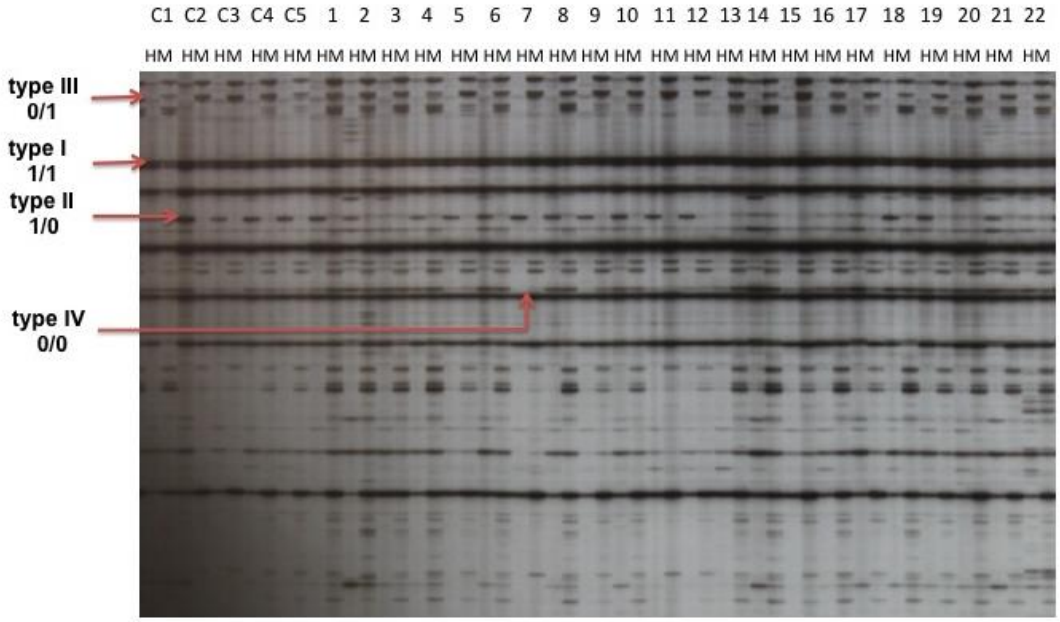


Figure 3.6 Example of MSAP profiles and classification of four types of bands.
MSAP banding pattern obtained using the primer combination *Eco*R I + ACA and *Hpa* II / *Msp* I + TAA.
C1-C5: Control lines. 1-22: Randomly selected hypomethylated lines. H: DNA subjected to *Eco*R I / *Hpa* II digestion. M: DNA subjected to *Eco*R I / *Msp* I digestion. The arrows indicate type I, type II, type III and type IV bands amplified.
The "1" represents the presence of bands and "0" represents the absence of bands for scoring purposes.

Table 3.6 Distribution of variant phenotypes among 5-azaC concentration classes.

	1.0mM	5.0mM	20mM	50mM
Early flowering lines	0	2	2	1
Late flowering lines	1	2	1	5
Large rosette diameter lines	1	0	0	0
Small rosette diameter lines	0	9	4	15

Table 3.7 A summary of DNA methylation profile in five grouped subpopulations after 5-azaC treatment.

Treatment (mM)	Type I (1/1)	Type II (1/0)	Type III (0/1)	Type IV (0/0)	Methylated cytosine %
0	110	11	73	52	38.37
1.0	112	11	78	46	36.67
5.0	114	9	74	51	37.2
20.0	117	8	77	44	35.04
50.0	129	11	70	35	30.77

Methylated cytosine (%) = [(II*2 + III*2 + IV*4) / ((I + II + III + IV)*4)]* 100
A total of 246 amplified loci were scored in every individual.

Table 3.8 Analysis of variance (ANOVA) comparing cytosine methylation levels among exposure to different concentrations of 5-azaC. Tukey multiple comparisons of means (95% family-wise confidence level).

Treatment	diff	lwr	upr	adj
1.0mM-0mM	-1.708	-7.8501	4.43405	0.92
5.0mM-0mM	-1.179	-7.6936	5.33563	0.98242
20.0mM-0mM	-3.2397	-8.9262	2.44672	0.45978
50.0mM-0mM	-6.4973	-12.378	-0.6168	0.02557*
5.0mM-1.0mM	0.529	-5.9856	7.04363	0.99919
20.0mM-1.0mM	-1.5317	-7.2182	4.15472	0.92806
50.0mM-1.0mM	-4.7893	-10.67	1.09123	0.14815
20.0mM-5.0mM	-2.0607	-8.1477	4.02625	0.85053
50.0mM-5.0mM	-5.3183	-11.587	0.95037	0.12289
50.0mM-20.0mM	-3.2576	-8.6606	2.14533	0.40455

*$p < 0.05$

in each 5'-CCGG-3' site in control revealed that the majority (54%, type III / (type II + type III + type IV), **Table 3.7**) occurs at a single cytosine base on each strand and 38% (type IV / (type II + type III + type IV), **Table 3.7**) were observed with methyl groups on all four cytosine bases. The remainder (8%, type II / (type II + type III + type IV), **Table 3.7**) of the methylation is found in a hemimethylated context.

The DNA methylation patterns observed among the epimutagenized population was explored using (PCoA). A total of 109 (44%) MSAP loci passed filtering and were designated as methylation-susceptible loci, these loci were used in a PCoA and the variances of the two largest components were plotted to describe the variation in both the control and 5-azaC treated groups (**Figure 3.7**). This analysis summarizes the highly dimensional MSAP data demonstrating that variation in DNA methylation patterns were observed among the sampled control lines and that the level of variation increased in the 5-azaC treated population. As anticipated, a large overlap was observed between the two populations, where the variation among the treated individuals expanded to occupy greater space encompassing the control lines, resulting in no significant difference between the two populations being detected by AMOVA (epigenetic distance $\Phi st = -0.006$, $p = 0.52$). The additional variation induced in the 5-azaC treated lines was revealed by a positive shift along the axis of both the first and second components, which explained 15% and 13% of the variance respectively. The number of MSAP banding patterns observed in each of the 27 individuals at each of the 246 loci was summarized and used to estimate percent methylation at each of the sampled cytosine bases (**Table 3.9**). The magnitude of DNA methylation change was correlated with 5-azaC concentrations, although variation among the control lines was observed. Examination of the strawberry epimutant population using MSAP markers indicated that changes in DNA methylation up to 10 fold beyond the background with a 2% range detected among control lines (37.2% to 39.8%) whereas the variation among the epimutant population ranged by 20% (19.1% to 39.2%). Curiously, although the observed changes in the epimutant population indicated a general depletion of DNA methylation there were individuals where DNA methylation increased. Interestingly, the line showing the
greatest reduction in DNA methylation (ERFv 153) was also present in an early flowering phenotype (**Figure 3.3 A**).

3.4.5 Inheritance of variant phenotypic traits

A total of five individuals were selected from the initial 5-azaC treated population (H4S8). The selected lines included the early flowering lines ERFv153 and ERFv148, the late flowering lines ERFv138 and ERFv141 and a line with a small rosette diameter ERFv65 (**Figure 3.8**). Progeny (H4S9) from each of these individuals were grown where flowering time and rosette diameter were measured. Transmission of the variant phenotype was observed from each of these selected lines to their progeny (**Table 3.10)**. The average flowering time of early flowering lines progeny was four days earlier and the late flowering lines progeny was 13 days later than the control lines. The average rosette diameter was 0.3 cm less than control lines. The extreme phenotypic values observed in the H4S9 generation were within treated families rather than within control families when the phenotypic scores were ranked. Although there was variation observed among the progeny, the family median had shifted and was significantly different from control lines. This was best exemplified in the progeny from the late flowering line ERFv141 where the distributions were almost distinct and average flowering time had become significantly later ($p < 0.0001$) (**Figure 3.9**).

3.4.6 Inheritance of DNA methylation patterns

The transmission of DNA methylation patterns through meiosis was assayed by subjecting the progeny from each of the four phenotypic classes, small rosette diameter (P1); early flowering (P2); late flowering (P3); and control (P4) to phenotypic evaluation and MSAP analysis. The H4S9 generation consisted of 27 individuals comprised of four individuals from the P1 class, eight individuals from the P2 class, seven individuals from the P3 class, and eight individual from the P4 class (**Figure 3.8**). MSAP analysis of the individuals from the H4S9 generation yielded a total of 333 loci where 43% were identified as methylation-susceptible loci. These data were subjected to multivariate analysis using PCoA to summarize the relationships among the individuals. The first two coordinates explained 28% of the total variance in DNA methylation with the first coordinate explaining 19% variance (**Figure 3.10**). As expected, the plot of the P4 class (control lines, H4S9) displayed a similar pattern, occupying the same area and shape, to the control lines examined in the previous generation (**Figure 3.7**). The MSAP data for the individuals in each of the phenotypic classes (P1-3) cluster together and separate themselves from the control population. The largest differences were observed between those individuals exhibiting an early flowering phenotype (P2) and the control population (P4), where the distance ($\Phi st = 0.1032$, $P < 0.0002$) separating the clusters was found to be significant. Pairwise comparisons showed that the methylation patterns in classes P1, P2 and P3 were significantly different from the control class (P4) (**Table 3.11**). The most significant

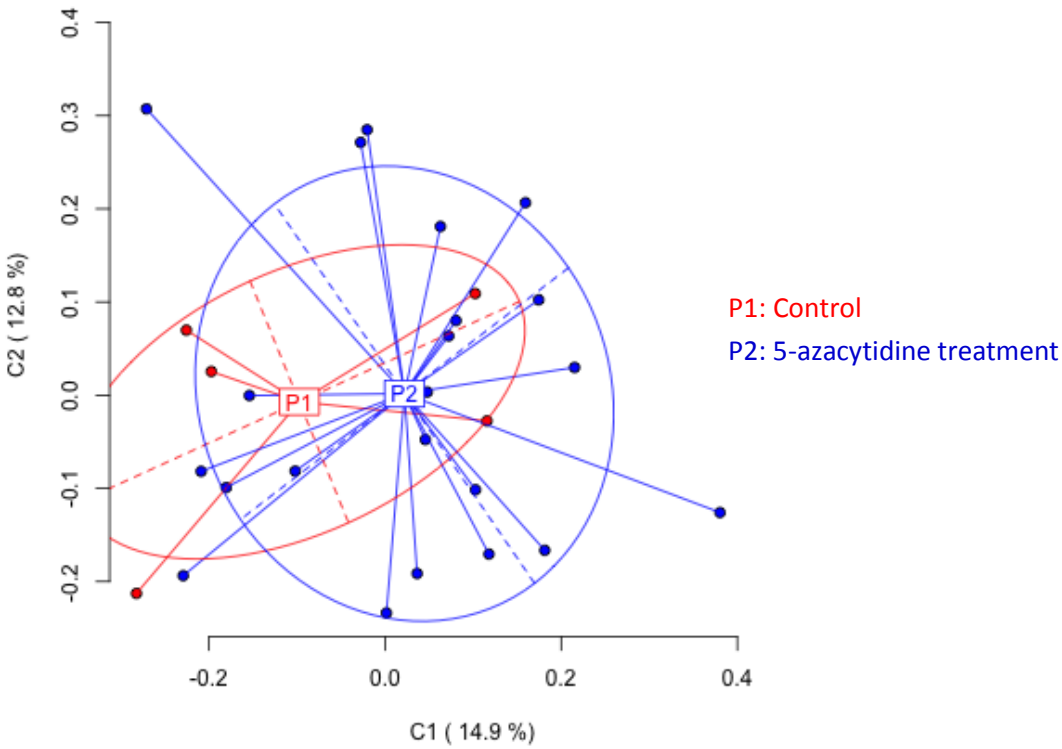


Figure 3.7 Principal Coordinates Analysis (PCoA) for DNA methylation differentiation between control lines and 5-azaC treatment lines using polymorphic methylation-susceptible loci (MSL) data.
The percentages in the first two coordinates (C1 and C2) show the contribution of variance (shown in brackets). Color-labeled P1, P2 are the centroids of the respective group. P1 represents five control lines and P2 represents 22 hypomethylated lines.

Table 3.9 A summary of DNA methylation profile in five control lines and 22 randomly selected epimutagenized population lines.

Number on gel	Lines	Treatment (mM)	Type I (1/1)	Type II (1/0)	Type III (0/1)	Type IV (0/0)	Methylated cytosine %	Relative DNA methylation %
1	Control 1	0	114	6	73	53	37.6	-0.77
2	Control 2	0	104	15	77	50	39.02	0.65
3	Control 3	0	108	10	70	58	39.84	1.46
4	Control 4	0	112	16	69	49	37.2	-1.18
5	Control 5	0	110	10	74	52	38.21	-0.16
6	ERFv 11	1	110	19	71	46	36.99	-1.38
7	ERFv 16	1	111	13	72	50	37.6	-0.77
8	ERFv 168	1	110	7	84	45	36.79	-1.59
9	ERFv 259	1	110	7	84	45	36.79	-1.59
10	ERFv 312	1	117	7	78	44	35.16	-3.21
11	ERFv 134	5	116	9	75	46	35.77	-2.6
12	ERFv 137	5	111	9	68	58	39.23	0.85
13	ERFv 157	5	114	10	75	47	36.38	-1.99
14	ERFv 345	5	113	6	76	51	37.4	-0.98
15	ERFv 45	20	116	12	63	55	37.6	-0.77
16	ERFv 141	20	115	15	74	42	35.16	-3.21
17	ERFv 147	20	118	10	77	41	34.35	-4.02
18	ERFv 217	20	111	6	92	37	34.96	-3.41
19	ERFv 295	20	122	5	75	44	34.15	-4.23
20	ERFv 328	20	120	6	78	42	34.15	-4.23
21	ERFv 329	20	121	12	63	50	35.57	-2.8
22	ERFv 95	50	123	11	71	41	33.33	-5.04
23	ERFv 127	50	120	12	73	41	33.94	-4.43
24	ERFv 132	50	113	14	78	41	35.37	-3.01
25	ERFv 140	50	111	10	76	49	37.4	-0.98
26	ERFv 153	50	165	8	60	13	19.11	-19.27
27	ERFv 246	50	126	12	70	38	32.11	-6.26
NA	Control lines mean	NA	110	11	73	52	38.37	0
NA	Epimutagenized lines mean	NA	118	10	74	44	34.96	-3.41

Methylated cytosine (%) = [(II*2 + III*2 + IV*4) / ((I + II + III + IV)*4)]* 100

Relative DNA methylation (%) = Methylated cytosine (%) in treatment lines - Methylated cytosine (%) in control lines.

A total of 246 amplified loci were scored in every individual.

Table 3.10 Phenotypic trait properties of progenies derived from selected lines.

	Early flowering (day)			Late flowering (day)			Small rosette diameter (cm)	
	ERFv148	ERFv153	Control	ERFv138	ERFv141	Control	ERFv65	Control
Mean	60.6	62.9	65.8	86.9	91.7	76.5	2.2	2.5
SD	5.6	7.2	6.9	3.0	6.5	7.1	0.7	0.7

SD = the standard deviation of the progeny from each selected line.

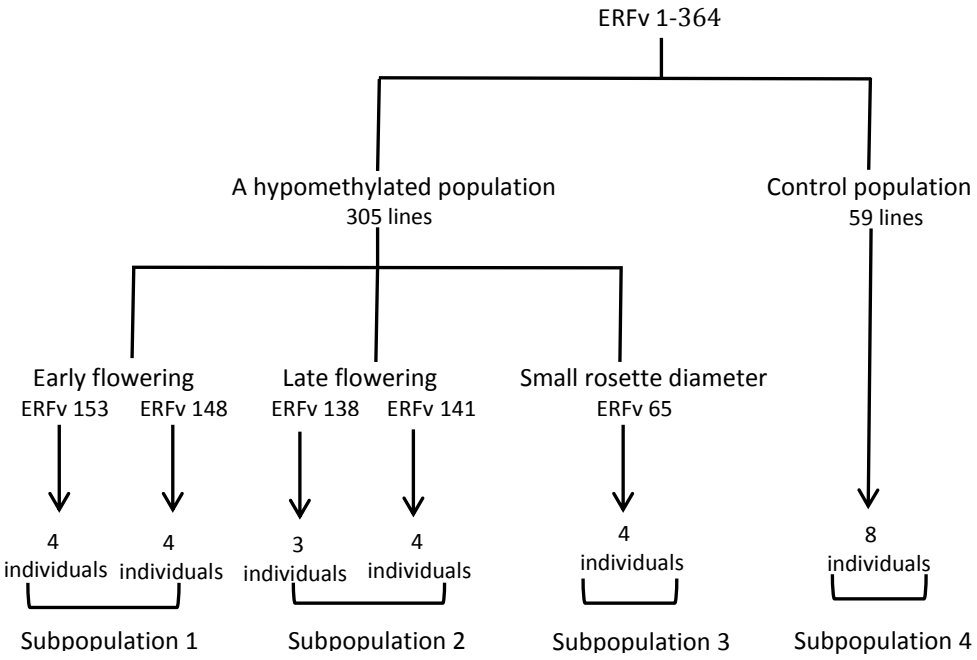


Figure 3.8 Overview of the hypomethylated population and subpopulation used in the MSAP study.
A set of 364 lines was obtained including 59 control lines and 305 epimutagenized lines. Four subpopulations consisting of 8 individuals of early flowering, 7 individuals of late flowering, 4 individuals of small rosette diameter and 8 individuals of control lines from the next generation were used in the MSAP study.

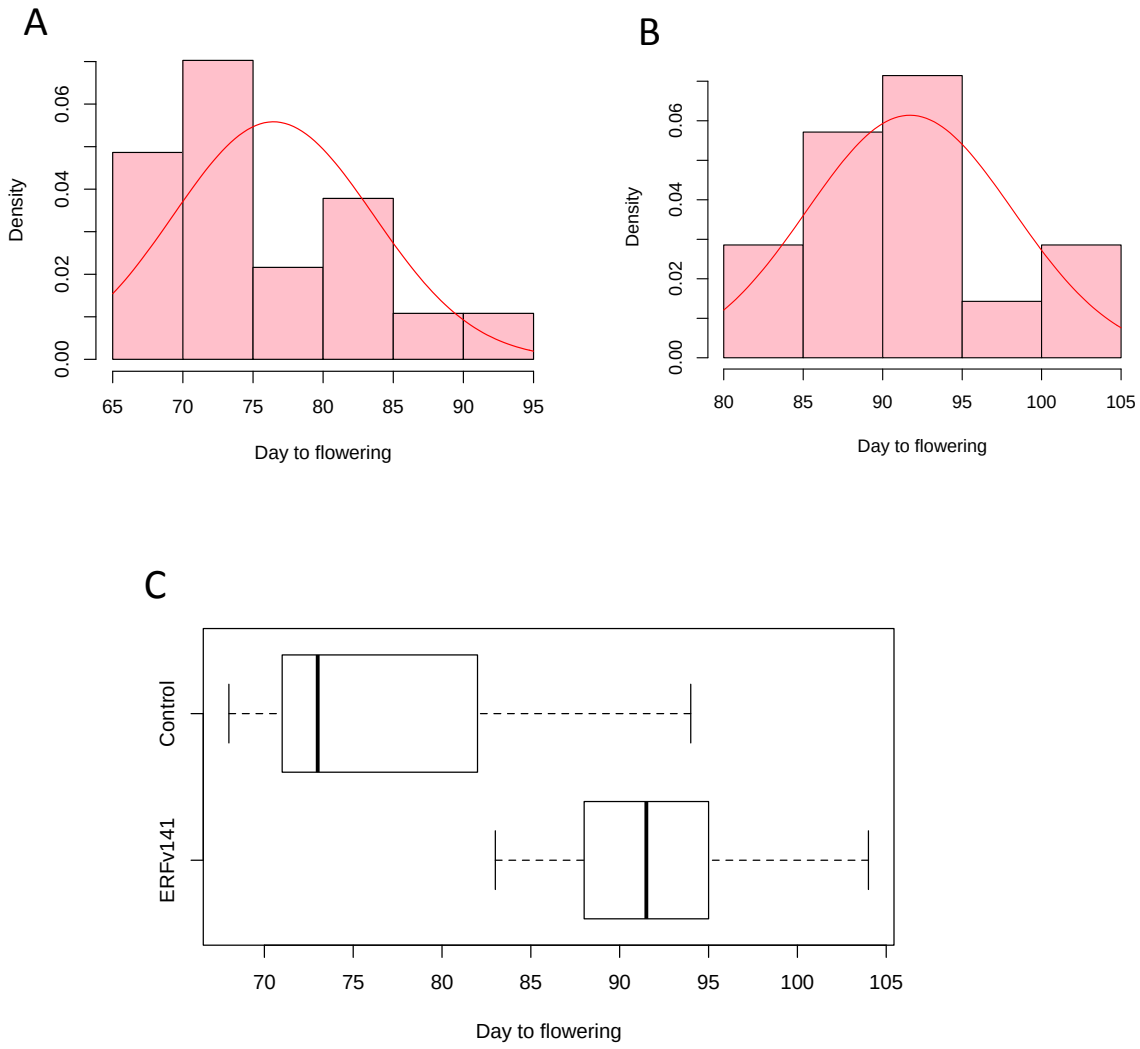


Figure 3.9 Transmission of flowering time in the late flowering line ERFv141 through meiosis.
A: Flowering time density histogram among the progeny of the control line. B: Flowering time density histogram among the progeny of the late flowering line ERFv141. C: Distribution of day to flowering displayed in boxplots among the progeny of control line and the progeny of the late flowering line ERFv141. The flowering time of the parental control and parental ERFv141 lines were 74 and 93 days, respectively.

difference was detected between the control and the early flowering population ($P < 0.0001$) and this difference is visualized by the formation of two distinct clusters in **Figure 3.10**. Interestingly, variegated chlorophyll pigmentation was observed in a single individual derived from the late flowering line ERFv134 (**Figure 3.3 B**), suggesting that novel phenotypes previously unobserved in the initial population might be revealed due to segregation of methylation patterns or other underlying factors in subsequent generations.

Multivariate statistical analysis of the MSAP data indicated enrichment of methylation patterns within families. This clustering suggests that DNA methylation patterns are heritable and is in agreement with data from other analyses (Morán and Pérez-Figueroa, 2011; Morán et al., 2013; Rico et al., 2014). Despite this, the sampled MSAP loci are, at best, linked to the factors underlying the variation and since they lack a genomic location they provide little utility beyond indicating that variation exists. In order to verify that DNA methylation patterns can be faithfully inherited from parents to offspring, methylation patterns were determined at defined loci. The loci were selected from an *in silico* analysis of the strawberry genome for enrichment in cytosine bases. These regions were randomly selected to assess the inheritance of methylation information and have no known bearing on the observed trait variation.

To test the fidelity of inheritance, a total of 21 lines were assayed including the same parental lines examined by MSAP, namely, the early flowering lines ERFv153 and ERFv148; the late flowering lines ERFv138 and ERFv141; the small rosette diameter line ERFv65 (**Figure 3.8**) and two control lines ERFv27, ERFv228 along with two siblings from the progeny of each selected parental line. Three target loci were selected from the *F. vesca* genome as being enriched for the presence of cytosine bases. The conversion efficiency of the sodium bisulfite treatment was adequate since each of cytosine bases were converted into uracil and sequenced as thymine when amplified from the Lambda genome spiked into each sample (**Figure 3.11 A**). The three target regions of the *F .vesca* genome sequenced after conversion with sodium bisulfite exhibited different levels of DNA methylation with 79% of the cytosine bases methylated at target site one (**Figure 3.11 B**), 59% of the cytosine bases methylated at target site two (**Figure 3.11 C**), and no methylation observed at target site three (**Figure 3.11 D**). Faithful inheritance of cytosine methylation in all three sequence contexts (CG, CHG, and CHH) was observed in target regions one and three in all of the 21 lines tested (**Figure 3.11 B, Figure 3.11 D**). At target region two, all methylated cytosine bases occurring in a CG

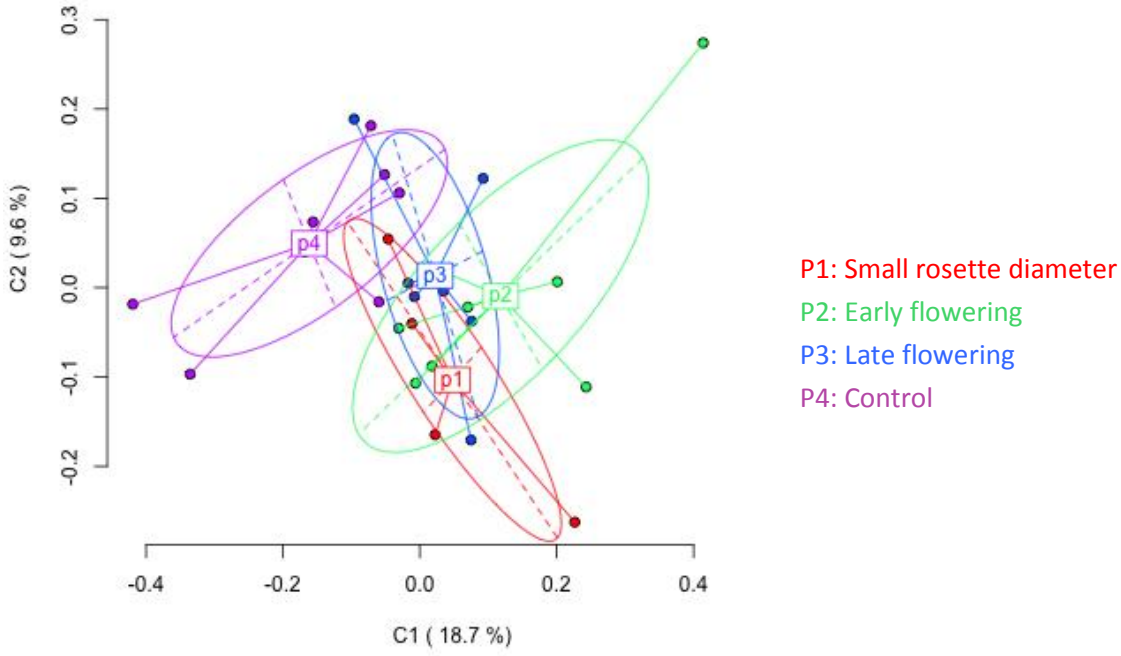
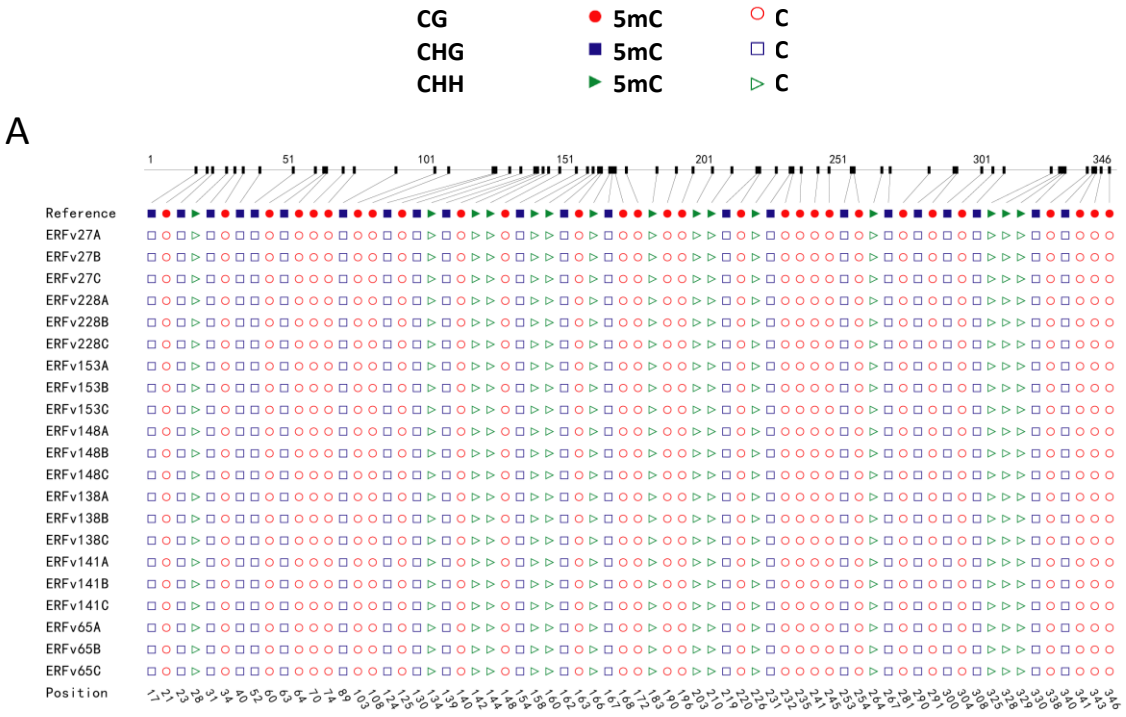
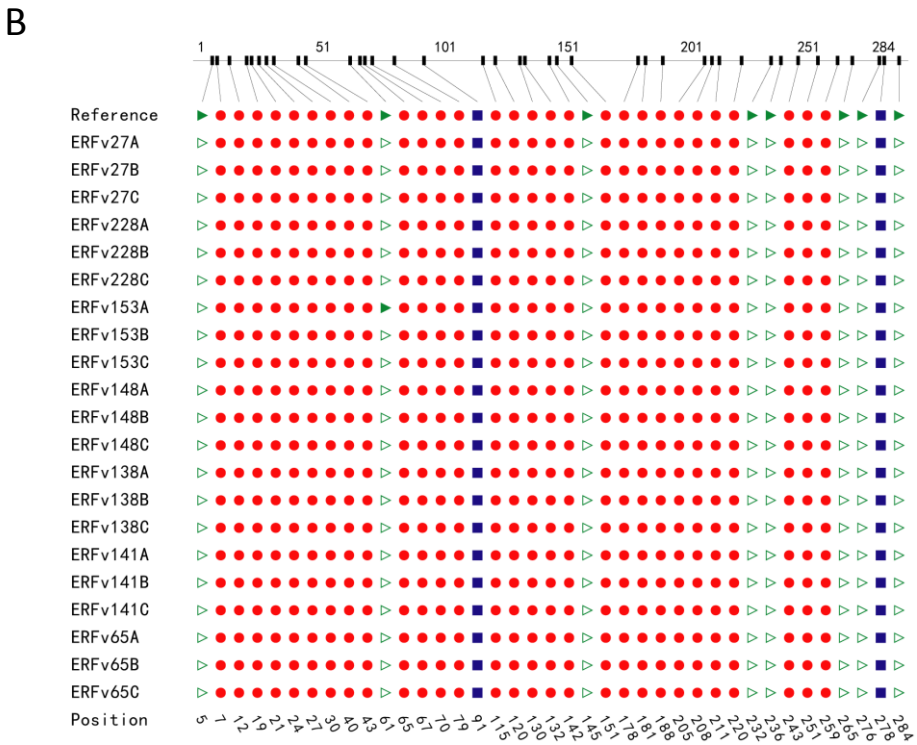


Figure 3.10 Principal Coordinates Analysis (PCoA) for DNA methylation differentiation between experimental groups using polymorphic methylation-susceptible loci (MSL) data. The percentages in the first two coordinates (C1 and C2) show the contribution of variance (brackets). Color-labeled P1, P2, P3, and P4 are the centroids of the respective group. P1: four small rosette diameter lines from line ERFv65. P2: four early flowering lines from line ERFv153, and four early flowering lines from line ERFv148. P3: three late flowering lines from line ERFv138, and four late flowering lines from line ERFv141. P4: eight control lines.

Table 3.11 Epigenetic differentiation between progenies of different phenotypic traits.

Samples pairs		Φst between samples	p value
Control (P4)	Small plant diameter (P1)	0.1288	0.0190
Control (P4)	Early flowering (P2)	0.1869	0.0001
Control (P4)	Late flowering (P3)	0.1174	0.0010
Late flowering (P3)	Small plant diameter (P1)	0.0308	0.2204
Late flowering (P3)	Early flowering (P2)	0.0356	0.0674
Early flowering (P2)	Small plant diameter (P1)	0.0144	0.2950


CG
CHG
CHH
5mC
5mC
5mC
C
C
C
A
Reference
ERFv27A
ERFv27B
ERFv27C
ERFv228A
ERFv228B
ERFv228C
ERFv153A
ERFv153B
ERFv153C
ERFv148A
ERFv148B
ERFv148C
ERFv138A
ERFv138B
ERFv138C
ERFv141A
ERFv141B
ERFv141C
ERFv65A
ERFv65B
ERFv65C
Position


B
Reference
ERFv27A
ERFv27B
ERFv27C
ERFv228A
ERFv228B
ERFv228C
ERFv153A
ERFv153B
ERFv153C
ERFv148A
ERFv148B
ERFv148C
ERFv138A
ERFv138B
ERFv138C
ERFv141A
ERFv141B
ERFv141C
ERFv65A
ERFv65B
ERFv65C
Position

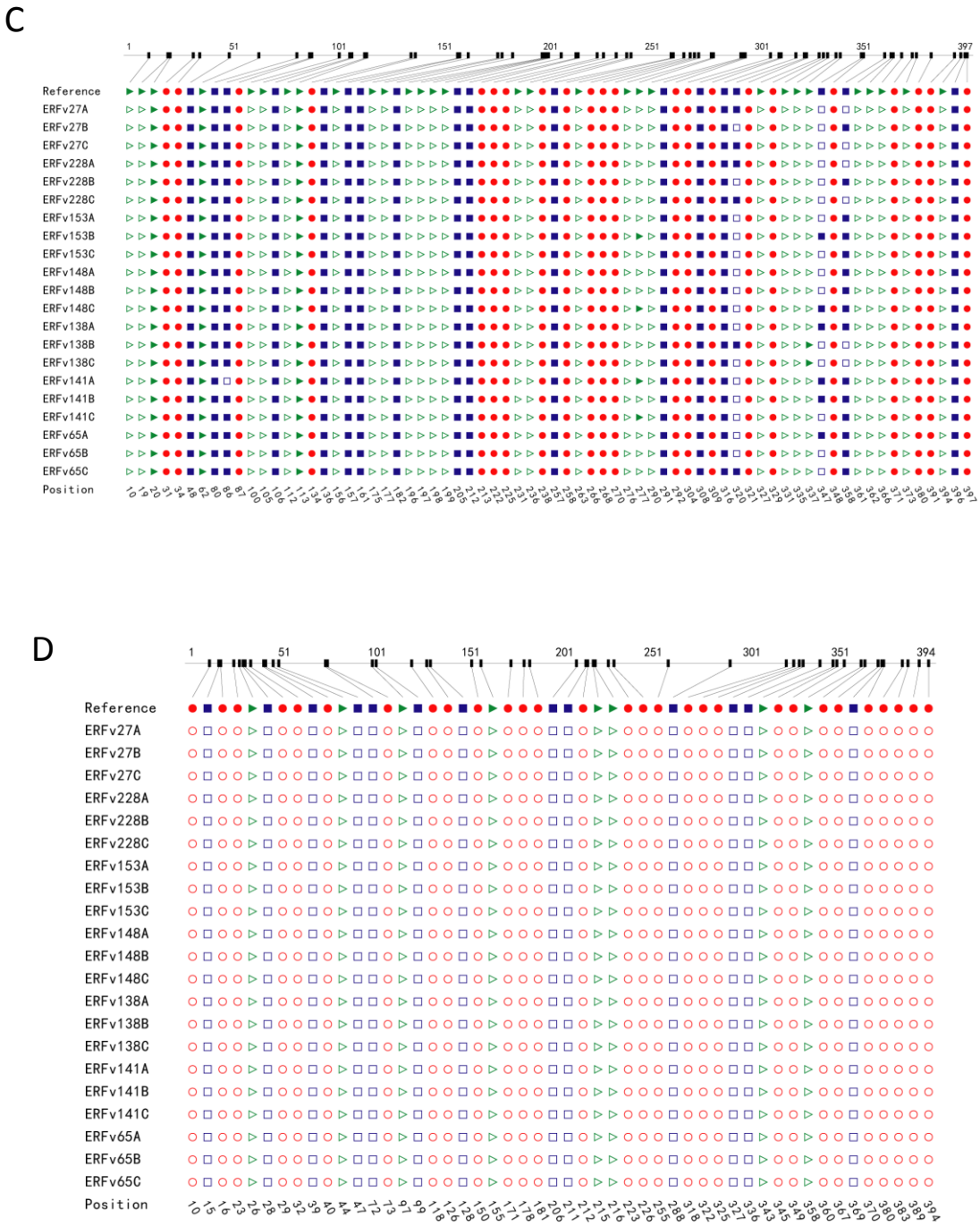


Figure 3.11 Methylation status of each cytosine in three CpG enriched strawberry target regions and the unmethylated Lambda control.
Assessment of bisulfite conversion efficiency using Lambda as an unmethylated control (A), target region one (B), target region two (C), and target region three (D). A total of 21 lines were assayed including control ERFv 27, ERFv 228, early flowering ERFv 153, ERFv 148, late flowering ERFv 138, ERFv 141, small rosette diameter ERFv 65 respectively. A represents the parent generation, B and C represent the next generation progeny derived from A. The Class 1, Class 2, and Class 3 represent CG, CHG, and CHH methylation, respectively. Filled shapes indicate methylated cytosines, and open shapes indicate non methylated cytosines. The number along the bottom indicates the cytosine position in the sequences.

context and the majority of methylated CHG and CHH (over 80%) were faithfully inherited from parent to progeny. However, variation in methylation pattern was observed at sequence positions 277, 320, 337, 347 and 358 (**Figure 3.11 C**). Overall, these results indicated that DNA methylation is faithfully transmitted through meiosis and this appears to occur with greater fidelity at CG positions.

3.5 Discussion

This study describes the generation of a novel hypomethylated population by treating seeds from a highly inbred *F. vesca* line with the cytidine analogue 5-azaC. It was demonstrated the population contains expanded variation for a range of quantitative characters and that some of these novel phenotypes and methylation patterns are transmitted through meiosis. This resource likely contains novel epialleles whose identification can lead to a more complete understanding of the cellular mechanisms that determine the extent of quantitative trait variation. This population is an ideal resource to address questions surrounding the extent that DNA methylation patterns exert over quantitative trait variation. The function and utility of epigenetic variation has been the subject of debate and intensive research (Becker and Weigel, 2012; Schmitz and Ecker, 2012). Studies have demonstrated the effect of epigenetic alterations on phenotypic variation taking advantage of the resources available in *Arabidopsis*, where EpiRIL derived from *met1* mutants show increased trait variation (Reinders et al., 2009). The phenotypic variation observed among the EpiRIL points to the potential of epigenetic variation for rapidly inducing new variation (Johannes et al., 2009; Reinders et al., 2009). Currently, however, strawberry does not have similar mutants to develop such resources. The use of pharmacological agents, such as 5-azaC, that inhibit DNA methylation is an alternative method of developing such resources. The overall effect of this treatment is the generation of a population of plants that are genetically near-identical yet possess unique, hypomethylated DNA methylation patterns. The one major function of DNA methylation is to silence repetitive elements although cytosine methylation occurs along the length of annotated genes (Zemach et al., 2010; Song et al., 2013). Also, alterations in DNA methylation hold the potential to change the regulation over transcription, altering the magnitude of gene expression and generating novel phenotypic variation.

Attributing phenotypic variation to epigenetic polymorphism can be confounded by the presence of genetic variation segregating in a population. Disentangling epigenetic and genetic contributions is challenging. To minimize the influence of genetic variation, the

hypomethylated *F. vesca* population described in this study was developed using a highly inbred genotype of Hawaii 4. The inbreeding through single-seed descent for eight generations would have resulted in >99% of the heterozygous alleles in the original Hawaii 4 ecotype to become homozygous. The high frequency of homozygous alleles was confirmed by the genomic AFLP marker analysis, which revealed no polymorphism. AFLP markers can simultaneously assay large numbers of loci throughout the genome without any prior sequence information (Mueller and Wolfenbarger, 1999). The results from this analysis and the history of the inbred *F. vesca* material indicated that the H4S8 population is genetically uniform. The detection of identical alleles at 219 loci across the seven chromosomes supports the premise that only *F. vesca* alleles derived from Hawaii 4 were present indicating that no alien *F. vesca* alleles had been introduced into the population through inadvertent hybridization. However, the possibility remained that SNP variation might be present among the hypomethylated lines and this could represent a significant source of variation should 5-azaC act as a mutagen in addition to reducing DNA methylation. Since the material used in this study was directly descended from the material used to generate the *F. vesca* genome sequence, the opportunity existed to estimate the potential of 5-azaC to behave as a mutagen and induce SNP variation. The relationship between the *F. vesca* material used in this study (H4S8) and the material used to generate the reference genome (H4S4) was an additional four generations of inbreeding. Whole genome sequence data were generated from seven individuals, four of which were exposed to high concentrations of 5-azaC (50 mM) and three from control lines (0 mM). These data were aligned to the reference genome sequence and SNP identified. The difference between the reference genome (H4S4) and the untreated H4S8 material indicates the level of SNP due to spontaneous changes. Polymorphisms between the reference genome and the 5-azaC treated individuals would include the same spontaneous mutation rate as the control lines in addition to any mutagenic effects of 5-azaC.

It is challenging to reliably identify loci with single nucleotide polymorphisms from alignments of short read sequence data to a draft reference genome sequence. In addition to obtaining adequate coverage of the genome, sufficient sequencing depth is required to discriminate actual polymorphisms from those generated by artifacts including short read sequencing and alignment errors that are compounded by errors inherent in the reference genome assembly. Using parameters to detect high quality SNP, 148 loci were detected as possessing an allele that might have occurred spontaneously. This number increased slightly to 153 loci when the 5-azaC lines were included, providing no evidence for appreciable

mutagenic activity from 5-azaC. The SNP data emphasize the similarity among the control material and lines treated to the highest concentrations of 5-azaC. A total of 1548 loci were detected with SNP revealing polymorphism among the control and 5-azaC treated lines, a number that was considerably less than the number of loci where the identical genotype was found in (29137) the control and 5-azaC treated lines yet differed to the reference allele. These data suggest that the individual line selected for sequencing was heterozygous at these loci or perhaps that the sequenced material was derived from more than a single individual. The SNP data indicate that the H4S8 material is heterozygous at a minimum of 1208 loci and is homozygous for alleles exclusive to the 5-azaC treated lines at 340 loci. Sequencing of additional control material might reduce this number. Nevertheless, 183 of these loci were found in annotated genes with 11 predicted to introduce a premiture stop codon or alter splicing and 119 to change the amino acid sequence of the translated protein. The genome sequencing data verified the near-identical genetic background of the *F. vesca* control and 5-azaC treated populations. Our observation is consistent with a previous analysis demonstrating that 5-azaC lacks signficant activity to change the primary DNA sequence and functions as an efficient inhibitor of DNA methylation (Holliday and Ho, 2002). These data do not rule out the presence of genetic polymorphisms from the 5-azaC population, as hypomethylation might result in an increase in transposition activity or other chromosomal rearrangements that are difficult to detect from short read alignment data to draft genome assemblies. However, the absence fragment length polymorphisms in the AFLP data suggest that these are rare events. The detection of any underlying alleles controlling the expanded trait variation will require segregation analysis to identify their origin and location.

As anticipated, the major effect of exposure to 5-azaC was the induction of DNA methylation changes and these were detected in the hypomethylated population through analysis using MSAP markers. The MSAP data revealed that the genome DNA of *F. vesca* appears to be predominantly unmethylated with 62% of the sampled cytosines lacking a methyl group. The estimation of hemimethylation among cytosine bases in *F. vesca* was placed at 8%. This figure is comparable to estimations made in related species such as rose (10%) and apple (6%) (Li et al., 2002; Xu et al., 2004). In our study 54% of methylation occurs at the internal cytosine on each strand and 38% were on all four cytosine bases, while the opposite was observed in rose and apple, in which the majority (55% and 70% respectively) of methylation occured on all four cytosine bases and lower frequencies (35% and 24% respectively) on the internal cytosine of each strand. Although this analysis provides some understanding of the *F.*

vesca epigenome, the MSAP data lack positional information complicating their interpretation and rendering them largely descriptive. Further, the MSAP analysis is restricted to examining methylation status only at 5'-CCGG-3' sites and these estimates are made using a subset of the potential CG loci. Thus, these estimates might contain some bias. Nevertheless, we utilized the ability of MSAP to rapidly determine the level of DNA methylation and showed that plants exposed to 5-azaC had a reduction in the level of methylation at the sampled loci. These data provide the evidence that the 5-azaC treatment induced epigenetic variation resulting in the development of a *F. vesca* hypomethylated population. As anticipated, methylation depletion followed a dose response where greater depletion was correlated with a higher 5-azaC concentration. The overall trend observed among the methylation classes described a shift in the frequency of type IV to type III and type I classes with the 5-azaC concentration increasing. However, only at the highest 5-azaC concentration did the frequencies change appreciably. Nevertheless, inhibition of DNA methylation resulting by 5-azaC is expected to generate novel differentially methylated regions (DMR) when compared to control plants, some of which might affect transcriptional activity leading to increased phenotypic variation.

Moderate dysregulation of gene expression can be exploited to alter quantitative characters. The quantitative traits that were focused on in this study included flowering time and rosette diameter, although other morphological and physiological changes were observed. Similar to the results obtained from the analysis of the *Arabidopsis* EpiRILs (Johannes et al., 2009) and a range of studies in other species using 5-azaC (Fieldes et al., 2005; Amoah et al., 2012), we observed a broader range of quantitative phenotypic variation among the hypomethylated lines when compared to the control population. The expanded variation observed for flowering time yielded plants of early and late flowering at both tails of the distribution. However, for rosette diameter only plants with a smaller rosette were observed, perhaps reflecting a lower growth rate. Similar data have been observed by others where dwarfed seedlings were found in rice (Akimoto et al., 2007). In both *Arabidopsis* and *Brassica rapa*, plants tended to flower late after exposure to 5-azaC (Bossdorf, 2010; Amoah et al., 2012), while in both flax and potato, flowering time variants following 5-azaC treatment reached anthesis significantly earlier than controls (Fieldes, 1994a; Marfil et al., 2012). The early flowering strawberry lines identified in this study did not appear stunted in their growth habit as what was observed in the flax early flowering variants, suggesting that these phenotypes are not merely the result of stress.

Transmission of the phenotypic variants through meiosis suggests that the phenotypic changes are due to heritable factors. This was exemplified by the progeny from the late flowering variant ERFv141 where >95% of the lines flowered significantly later than the control population. It was also observed that the early flowering and small rosette diameter were transmitted to the following generation although the distributions were less distinct than for the ERFv141 material. Although the underlying factors controlling this variation remain undetermined, future characterization of these lines could uncover the factors responsible. MSAP is ill suited for identifying the underlying factors, due to the low number of sites sampled, restriction site distribution and the low resolution of methylation information provided. However, accessibility of this technology makes it ideal for initial characterization and assessment of variation in methylation patterns. Analysis of the H4S9 progeny derived from the selected lines possessing early flowering, late flowering, small rosette and control phenotypes revealed different DNA methylation patterns. Clustering of MSAP patterns from phenotypically related individuals suggests that some loci might be associated with the factors controlling these characters. The distances separating these clusters were demonstrated to be statistically significant. This was especially pronounced for the early flowering phenotype. The early flowering line ERFv153, which was treated with 50 mM 5-azaC, was the most hypomethylated with the methylation level at sampled loci reduced by 19.1%. Clustering of the H4S9 generation by phenotypes (**Figure 3.10**) contrasts with the unrelated pattern observed among the individuals of H4S8 generation (**Figure 3.7**) where the changes occurred in all directions in each dimension of the data summarized using eigenvalues. These data suggest that some of the DNA methylation patterns are associated, possibly through linkage, with the underlying heritable factors but are themselves unlikely to be directly related.

Comparing the inheritance of DNA methylation patterns using MSAP data is complicated by potential variation in PCR amplification and resolution of fragments under differing electrophoresis conditions. To overcome these limitations, DNA methylation patterns at three different target loci were amplified after conversion using sodium bisulfite. This approach ensured direct comparison of the methylation status of the same cytosine bases from 21 lines from which the fidelity of inheritance was estimated from the transmission of marks from each of the seven parents to two progeny lines. The three target regions are unrelated to any trait, but the methylation patterns are faithfully inherited for majority of the cytosine bases analysed. Inheritance was found in target region one and three in each sequence context.

Inheritance was observed at CG sites in target region two and only slight variation was observed in CHG and CHH contexts. This variation might result from ncRNA-directed *de novo* methylation, sequencing errors or polymorphism to the reference genome sequence. Nevertheless, these data indicate faithful transmission of DNA methylation patterns from parent to progeny at the majority of methylation sites particularly CG sites.

Flowering time is an important multigenic trait commonly targeted in plant breeding programs. In a situation where the introgression of alleles from wild species can compromise quality traits, the use of compounds such as 5-azaC to induce desirable novel variation without changing the combination of alleles selected by breeders could play a part in future crop improvement strategies. This study establishes a resource containing expanded phenotypic changes, which were transmitted through meiosis and can potentially be subjected to selection, demonstrating the potential of this approach. Although this resource was established in highly inbred material and whole genome DNA sequence analysis indicated that exposure to 5-azaC did not induce new SNP, genetic variation such as chromosomal rearrangements might have resulted from the induced hypomethylation. However, the frequency of their occurance is low since no fragment length variation was observed using AFLP. Further research is required to identify the factors, underlying these phenotypic traits. The material described makes an ideal resource to examine the heritability of this variation through multiple generations and upon the identification of the underlying factors, and the potential of drugs such as 5-azaC for generating useful variation.

3.6 Conclusions

This study describes the generation of a population exhibiting expanded variation in strawberry. *F. vesca* is increasingly being used as a model plant for the Rosaceae family and its clonal and sexual strategies offer new experimental opportunities to evaluate the transmission of DNA methylation marks. We confirmed that the Hawaii 4 material used for population development was genetically uniform and that no appreciable increase in SNP frequency was induced by exposure to 5aza-C. Increased variation in DNA methylation profiles and phenotypic variation for several quantitative traits was observed. Moreover, the value of this resource was increased with the demonstration that the variant traits and DNA methylation patterns could be transmitted through meiosis. However, further work was required to detect if the variant traits were heritable over generations and subject to selection as shown in Chapter 4.

4.0 STABLE EPIGENETIC VARIANTS SELECTED FROM AN INDUCED HYPOMETHYLATED *FRAGARIA VESCA* POPULATION

4.1 Abstract

Epigenetic inheritance was transmitted over five generations through selection of extreme early, but not late flowering time phenotypic lines in *F. vesca*. Epigenetic variation was initially artificially induced by DNA demethylation using 5-azacytidine. This provides a new approach for studying the role of epigenetics on complex quantitative trait improvement in plants, as well as expediting potential new horticulture cultivar release through either seed or vegetative propagation. It is the first report to explore epigenetic variant selection and different phenotypic trait inheritance in strawberry. *F. vesca* is a diploid strawberry plant in the Rosaceae family, the family containing the most important horticulture temperature fruit crops including apples, cherries, peaches, pears, plums as well as ornamental plants such as roses. *F. vesca* is a useful model system to examine the transmission of phenotypic traits induced by epigenetic variation due to its short 3.5 month seed to seed life cycle in the greenhouse and modified stems (stolons) to clonally propagate daughter plants.

The *F. vesca* material was used to assess if selection can be applied and transmission frequency of these traits were determined across generations. The early flowering and late stolon traits were successfully transmitted through meiosis for at least five and three generations respectively, as demonstrated in the Early Flowering line (EF4) and the Late Stolon line (LS). This study demonstrated the stability of the early flowering phenotype based on selection of variants from a hypomethylated population over at least five generations. Stable mitotic transmission of the early flowering phenotype was also demonstrated using clonal daughters derived from the 4th Generation (S4) mother plant. On average, the early flowering families flowered three to five days earlier than the control families when transmitted through meiosis and six to ten days earlier when transmitted through mitosis. This study provides evidence of heritable selection of phenotypic traits based on 5-azaC induced DNA methylation in strawberry through both meiosis and mitosis, which is meaningful to both breeding programs and commercial horticulture.

4.2 Introduction

Epigenetic marks may be passed down from parent to offspring without altering the primary

DNA sequence. DNA methylation, histone modification, and small RNA interference are all epigenetic marks. DNA methylation patterns are the most heritable marks and can be associated with patterns of histone modifications, regulating gene transcription (Feng and Jacobsen, 2011). Additionally, small RNAs are also involved in the induction of DNA methylation (RNA directed DNA methylation) and histone modification processes (Law and Jacobsen, 2010). The composition of these marks together affect the conformation, topology of chromatin, regulation of gene expression, controlling growth and developmental transitions in plants (Hirsch et al., 2012; Verkest et al., 2015). When these modifications interact with the protein coding genes in the genome, alterations to phenotypic traits can be generated (Becker and Weigel, 2012). These phenotypic characteristics may be of significant commercial value but incorporation of valuable traits through traditional breeding approaches takes considerable time.

In *Arabidopsis*, continuous distribution and variation was observed in complex quantitative traits such as flowering time, plant height and plant growth in epigenetic Recombinant Inbred Lines (epiRILs) having similar genome but varying DNA methylation levels (Johannes et al., 2009; Reinders et al., 2009). DNA methylation status was further demonstrated to be transmitted from parent to offspring through both mitosis and meiosis in *Arabidopsis* (Kakutani et al., 1999). The mitotic inheritance is referred to as the propagation of epigenetic modifications through somatic cell division (Martin and Zhang, 2007). Transgenerational meiotic inheritance of quantitative traits such as flowering time and plant height were also reported in families of epiRILs in *Arabidopsis* (Johannes et al., 2009; Zhang et al., 2013). In *Brassica*, the transgenerational inheritance of seed size, plant stature and floral morphology has been examined using 5-azaC (Amoah et al., 2012). The Rosaceae family members such as strawberry (*F. vesca*, *F.* ×*ananassa*), are commercially propagated through mitotic generations (clonal "daughters") as well as meiotically (achenes, seeds) for breeding purposes. However, selection of stable epigenetic mutants in strawberry or other Rosaceae family members has not yet been investigated.

Not all epigenetic marks are transgenerationally heritable, with many being reset at meiosis which likely cause the subsequent disappearance of phenotypic traits induced by epigenetic changes in the following generations (Hauben et al., 2009; Falke et al., 2013). In the *Arabidopsis* epiRILs study, after two to five generations, DNA remethylation occurred in

some of the hypomethylated variants (Johannes et al., 2009). In *Brassica napus* lines, the acquired epigenetic component disappeared in the progenies and the phenotypic trait of lower respiration rate returned to the level of control (Hauben et al., 2009). In the stress-induced *Arabidopsis* lines possessing epigenetic memory, the modified epigenome can be transmitted to the next generation, but it disappeared after two generations without stress treatment (Suter and Widmer, 2013).

Nevertheless, inherited epialleles are possible under selection even though these epigenetic patterns may be transient. Underlying the transgenerational high heritability in flowering time and plant height, stable epialleles across the genome play an important role in this process (Johannes et al., 2008; Johannes et al., 2009). The generation of epiallelic variation and the resultant phenotypic diversity in seed size and composition that allowed selection was demonstrated in *Brassica rapa* inbred lines (Amoah et al., 2012). In *Arabidopsis,* a quantitative genetics approach examined the heritability of flowering time, plant height, total biomass and root:shoot ratio in different environmental conditions which showed the action of applied selection on phenotypic plasticity (Zhang et al., 2013). The artificial selection of extreme lines which might contain the phenotypic-related epialleles were reported in *Brassica napus*, the selected lines were self-fertilized and families were generated and evaluated for energy use efficiency (EUE). Following rounds of recurrent selection, higher EUE was successfully inherited through the generations and were distinguished based on epigenetic (DNA methylation and histone modification) status (Hauben et al., 2009). Therefore, recurrent selection for certain specific phenotypic traits associated with epialleles such as EUE may enable breeders to fix those traits in crops of commercial value.

We used the diploid *F. vesca* lines identified from the hypomethylated population generated in Chapter 3 to investigate the heritability of variant traits including early and late flowering time, and late stolon emergence time. The question of whether the lines with specific phenotypic traits possessing varied epigenetic status can be subject to selection was explored in this study. Since *F. vesca* can propagate either through generative or vegetative processes, it is an interesting model system to test the transmission of flowering time through both meiosis and mitosis from the same mother plants. This is the first report to explore epigenetic variants selection and different phenotypic traits inheritance in strawberry and presents a new approach to accelerate breeding by inducing heritable epigenetic lines of horticultural importance through 5-azaC application.

4.3 Materials and Methods

4.3.1 Plant materials

The plant materials examined in this chapter represent successive generations of control and hypomethylated families of *F. vesca* derived from material described previously (Chapter 3). The original generation (H4S8) used to produce the hypomethylated population is referred to as S0. Fifty-nine of these lines were control (no 5-azaC treatment) and the remaining 305 lines were treated with 5-azaC. These lines are a direct descendent of the plant (H4S4) *F. vesca* on which the reference genome was sequenced (Shulaev et al., 2011). We confirmed that the generation S0 is genetically identical but epigenetically unrelated since plants possesses expanded variation in DNA methylation caused by a loss of methylation at random loci at the sites of 5-azaC incorporation, following a dose-response function. This additional epigenetic variation results in expanded phenotypic variation following a hidden stochastic function. The hypomethylated individuals in subsequent generations (S1 through S5) are related, as families derived from selfing.

To follow the pedigree of the plant material throughout the experiment, all the 5-azaC treated lines and control lines were assigned a sub-code to the original names provided to each line (ERFv1 to ERFv364) described in Chapter 3. All *F. vesca* plants grown in this study followed the same procedure: dried achenes (seeds) were placed on wet filter paper in petri dishes and allowed to imbibe and germinate before being transplanted to soil and grown in the greenhouse at 23 ± 2°C day and 18 ± 2°C night with 18 / 6 h day / night photoperiod.

Pollination bags were placed over opening strawberry flowers for seven days to ensure the next generation of achenes was derived from self-pollination. The achenes were harvested from mature berries, dried under low heat using a food dehydrator before being harvested and maintained separately. The S-suffix number denotes each successive generation of selfing, where the generations S0 through S5 represent the inbred generations H4S8 through H4S13.

4.3.2 Phenotypic assessment of quantitative traits

The scoring of flowering time was performed as described in Chapter 3; stolon emergence time was scored as the number of days from sowing to the time when the first stolon was observed on the mother plant. Plant cultural conditions were as described in Chapter 3. To

account for differences observed among repeated measurements conducted across time and seasons, a control population was grown at each generation and all plants were randomized on the greenhouse bench.

4.3.3 Artificial Selection for the desired traits

A strict pedigree method was used based on families, and a summary of the number of lines assayed per family is presented (**Figure 4.1**)**.** Seeds were collected from the selected S0 plants and were grown to maturity to produce the subsequent generation 1 (S1). The phenotypic traits of the S1 population were recorded. The most extreme individuals within the S1 generation were selected to produce generation 2 (S2), rather than the family means. The selection based on the individuals within the most extreme family was used in S2, and the next rounds of selection in the subsequent generations followed the same rule. A total of five generations (S1, S2, S3, S4, S5) were selected for early flowering time, three generations (S1, S2, S3) were selected for stolon emergence time, and two generations (S1, S2) were selected for late flowering time.

In the evaluation of transgenerational inheritance of flowering time through mitosis, three daughter plants of the S5 generation derived from the first three stolons of each of five mothers studied were transferred into new four-inch pots when the width of the first triple leaves was 3cm as shown in **Figure 4.2**. A total of 15 plants were derived from three S4 generation families (EF4-9-13-15, EF4-14-7-6, EF4-14-7-14) with three daughter plants (three replications) and the average flowering time used for data analysis. Flowering time was scored as the number of days from transplanting daughter plants to anthesis determined by the opening of the primary flower of the inflorescence.

4.3.4 Statistical analysis

Phenotypic measurements made in the hypomethylated S0 generation were ranked to identify the lines expressing the extreme phenotypes. The mean and variance of each trait measured in the control S0 generation was determined by the Z-test to evaluate if each hypomethylated S0 individual was significantly different from the control mean. In the following generations, they were subjected to statistical analysis using two-sample Student’s t-test. Statistical analysis on the data of each trait measured used a 95% confidence interval to select hypomethylated individuals as potentially belonging to a population that differs from the

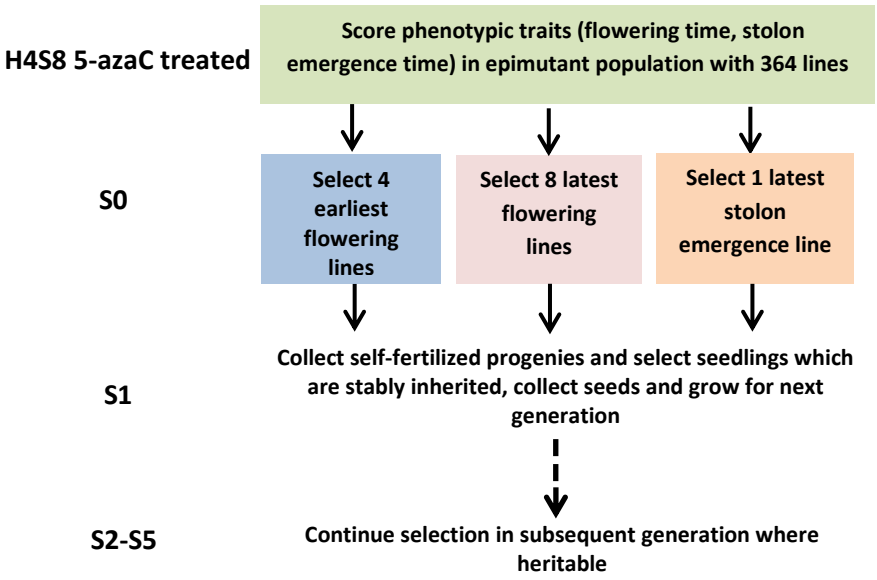


Figure 4.1 Selection for quantitative traits across seed generations in *F. vesca*.
The selection was started with the hypomethylated seedling population with genetically identical background derived from 5-azacytidine treated seeds.

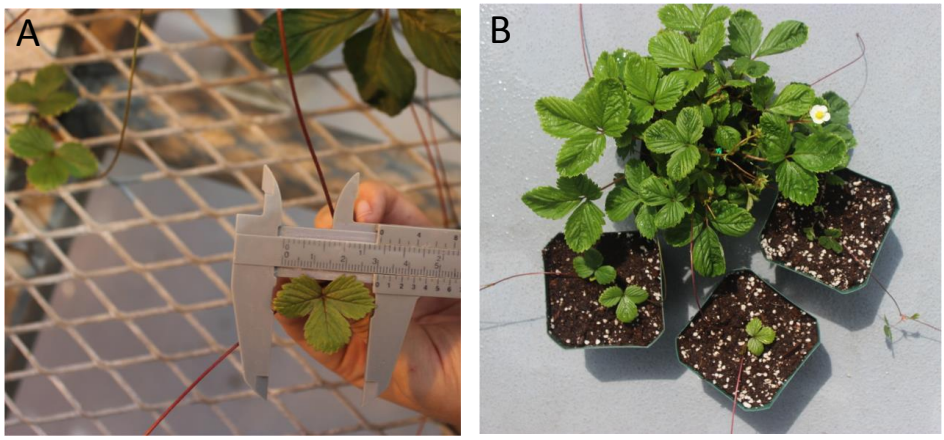


Figure 4.2 Example of clonal daughter plants derived from vegetative stolons in *F. vesca*.
A: The first triple leaves (size 3cm) from the primary daughter plant; B: Three daughter plants from the same mother plant were transferred into new pots.

control population. Comparisons between hypomethylated lines and control lines were restricted to those grown at the same time. All data was analyzed using statistical software R.

In order to establish a standard to compare the proportion of early and late flowering lines in the epimutant and control population, we defined the lines with flowering time less than the mean of control minus one standard deviation (SD) (mean-SD) as early flowering lines, and similarly the lines with flowering time greater than the control plus one standard deviation (mean+SD) as late flowering lines. A similar process was applied to the late stolon emergence lines. We defined the lines with stolon emergence time greater than the control plus one standard deviation (mean+SD) as late stolon emergence time lines. The equality of proportions of early flowering, late flowering, and late stolon emergence time in epimutant and control lines in each generation was evaluated by Chi-square.

4.4 Results

4.4.1 Initial selection of lines with extreme phenotypic traits (S0 and S1 generations)

The S0 generation was used to initially select extreme phenotypes of early and late flowering and late stolon emergence. In total, four early flowering lines (ERFv148, ERFv157, ERFv168, ERFv153, renamed EF1 to EF4 respectively), eight late flowering lines (ERFv16, ERFv132, ERFv127, ERFv131, ERFv134, ERFv140, ERFv141, ERFv138, renamed LF1 to LF8 respectively) (see Chapter 3 for the distribution of the population), and a single line expressing late stolon (LS) development (**Figure 4.3**) were selected as possessing the most extreme phenotype that significantly differed from the control population. ERFv153 (EF4) was selected in Chapter 3 for DNA methylation and DNA sequencing analysis, in which the greatest changes in DNA methylation level but no primary DNA sequence changes in genes controlling flowering time were detected. The early flowering sample sizes in each family of the S1 generation varied. The smallest population comprising only seven individuals and the largest comprising 22 individuals, while the control family consisted of 36 individuals (**Table 4.1**). This variation in family size was due to a combination of seed availability or plant survival after transplantation. The distribution of flowering time was already reported in Chapter 3 **Figure 3.6A**. The mean and standard deviation for flowering time distribution of the control population in the S1 generation was 65.8 and 6.9 days, respectively. By contrast, the flowering time mean values for the families derived from hypomethylated lines ranged from 60.6 to 66.6 days with standard deviations ranging from 5.6 to 8.3 days (**Table 4.1**). The

greatest significant difference was observed between the control and EF1 families where a difference of 5 days separated the means. Furthermore, a Chi-square test was used to determine the equality of proportions of early flowering time in early mutant and control lines. Although no difference was detected in the proportion of early flowering and control lines in families of the S0 generation, a significant difference in the S1 generation was detected ($p = 0.03$) (**Table 4.2**). In the S1 generation, the percentage of early flowering individuals (33.3%) in the early flowering epimutant lines was more than twice the percentage (13.9%) of control lines. When all families in the S1 generation based on phenotypic traits were combined (**Figure 4.4 A**), the proportion of early flowering individuals was higher in the early flowering epimutant lines than in control lines.

The transmission of the late flowering trait focused on eight families with population sizes ranging between seven and 26, these were compared to a control family consisting of 59 individuals. The control family mean was 76.5 days to flowering with a standard deviation of 7.1 days. The mean and standard deviation values describing the distributions of the hypomethylated families ranged from 73.1 ± 8.4 to 91.7 ± 6.5 days. The greatest difference between control and hypomethylated family means was around 14 days. Similar to the observation made on the early flower lines, when all hypomethylated lines were combined, there were individuals at the extremes of the entire distribution from families LF7 and LF8 where their flowering time family mean also was significantly late compared to the control mean value. Similar to early flowering lines, although no difference was detected in the proportion of late flowering and control lines in families of the S0 generation, a significant difference ($p = 0.01$) of late flowering individuals in the late flowering epimutant lines was also observed when compared to control lines in the S1 generation, with the percentage of late flowering lines more than twice the control in the S1 generation (**Table 4.3**). When all families in the S1 generation based on phenotypic traits were combined (**Figure 4.4 B**), the proportion of late flowering individuals was higher in the late flowering epimutant lines than in control lines.

A total of 38 out of 285 lines showed significantly late stolon emergence time (**Figure 4.3**). The significance of stolon emergence time was determined using the Z test with the distribution shown in **Figure 4.3.** One extremely late stolon emergence line named LS was selected to study the transmission of stolon emergence time. The phenotypic assessment for stolon emergence revealed data that are positively skewed, differing significantly from a

Table 4.1 Phenotypic trait properties of selected lines across different seed generations.

Phenotypic trait	S1 generation				S2 generation				S3 generation				S4 generation			
	Families	Mean	SD	Size	Families	Mean	SD	Size	Families	Mean	SD	Size	Families	Mean	SD	Size
Early flowering	EF1*	60.6^{+}	5.6	21	EF1-3	75.3^{+}	3.8	12	EF4-9-6*	68.7^{+}	3.5	20	EF4-9-13-15*	67.8^{+}	3.3	22
	EF2	66.5	5.9	22	EF1-5	78.5	5.6	15	EF4-9-10*	69.8	3.1	18	EF4-14-7-6*	69.4	3.1	27
	EF3	66.6	8.3	7	EF1-10	79.5	6.9	22	EF4-9-13*	69.2	1.7	18	EF4-14-7-14*	68.6	2.9	30
	EF4	62.9	7.2	19	EF1-20	79.8	7.3	19	EF4-14-3*	69.8	3.9	18	Control	73.2	4.7	20
	Control	65.8	6.9	36	EF4-9*	71.6	2.5	16	EF4-14-7*	67.8	1.9	20				
					EF4-14*	72.3	2.7	17	Control	72.7	2.2	45				
					EF4-18	75.8	2.8	21								
					Control	77.5	5.9	60								
Late flowering	LF1*	88.2	9.6	19	LF7-3	79.5	7	15								
	LF2	73.1	8.4	18	LF7-4	80.5	7.1	8								
	LF3	80.3	4	12	LF7-6	78.4	6.6	30								
	LF4	81.9	2.3	11	LF7-7	79	6.7	8								
	LF5*	83.6	2.3	26	LF8-1	76.8	9.8	4								
	LF6*	87.9	9.7	7	LF8-5	78.8	6.7	11								
	LF7*	91.7	6.5	14	LF8-7	79.5	6.3	23								
	LF8*	86.7	3	7	Control	79.4	7.1	28								
	Control	76.5	7.1	37												
Late stolon emergence	LS	59.1$^{\#}$	8.6	55	LS-54	56.1$^{\#}$	9.9	32	LS-56-10*	61$^{\#}$	12.3	24				
	Control	61.4	5.6	55	LS-56*	65.5	17.6	28	LS-56-25*	69	8.6	28				
					Control	58.5	10.3	32	Control	45.2	7.5	45				

*p <0.05. “Size” = the number of individuals. + day to the first flowering. # day to the first stolon emergence.

Table 4.2 Composition of the early flowering lines in each seed generation in epimutant and control lines.

Generation	Early flowering epimutant			Control			χ^2	p value
	Number of Early flowering lines	Number of lines	%	Number of Early flowering lines	Number of lines	%		
S0	65	293	22.2	14	59	23.7	0.07	0.80
S1	23	69	33.3	5	36	13.9	4.57	0.03
S2	14	61	22.9	4	30	13.6	1.17	0.28
S3	70	94	74.5	6	45	13.3	45.90	<0.001
S4	43	79	54.4	4	20	20	7.59	0.01
S5	18	36	50	1	24	4.2	13.98	<0.001

Early flowering lines were determined by the number of lines with flowering time less than the mean of the control minus one standard deviation (mean-SD).

Table 4.3 Composition of the late flowering lines in each seed generation in epimutant and control lines.

Generation	Late flowering epimutant			Control			χ^2	p value
	Number of Late flowering lines	Number of lines	%	Number of Late flowering lines	Number of lines	%		
S0	33	293	11.3	10	59	17	1.48	0.22
S1	38	107	35.5	9	59	15.3	7.69	0.01
S2	18	99	18.2	7	28	25	0.64	0.42

Late flowering lines were determined by the number of lines with flowering time greater than the mean of the control plus one standard deviation (mean+SD).

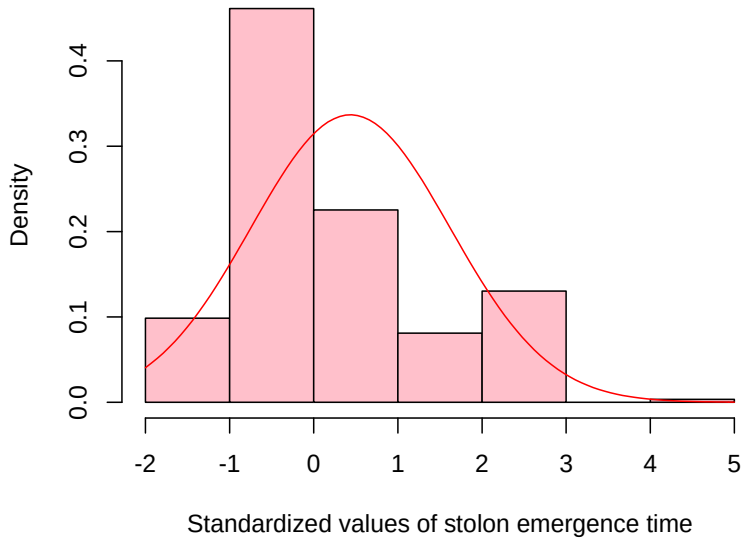
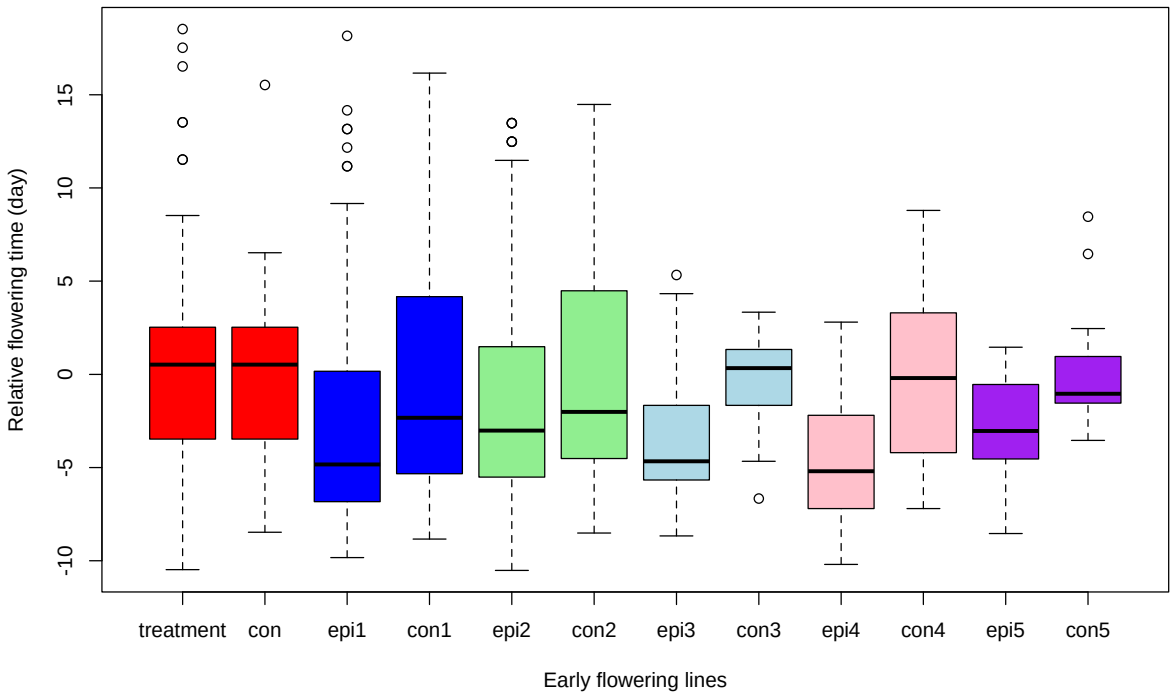
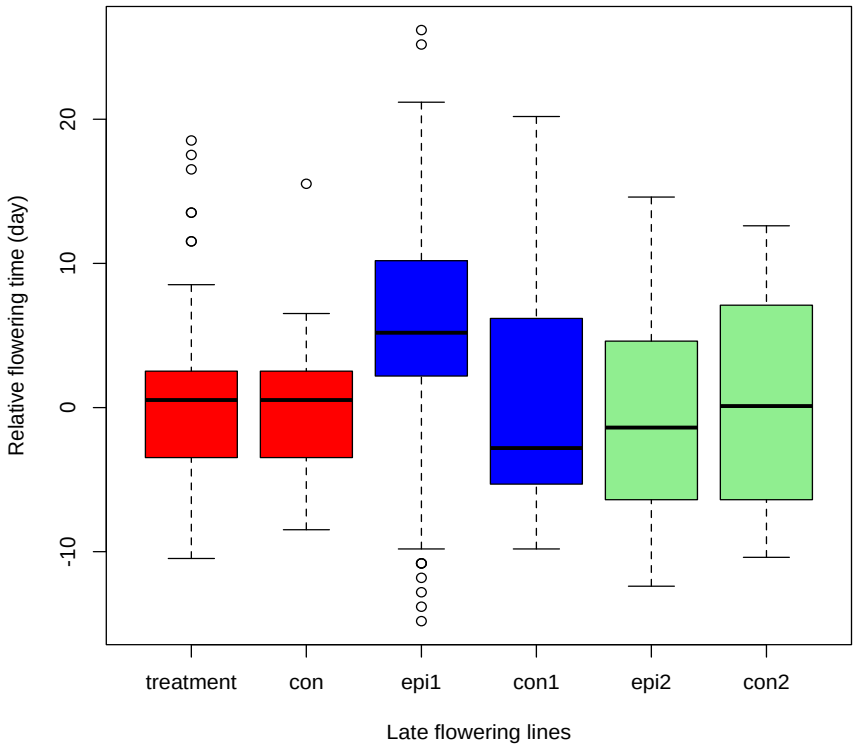


Figure 4.3 Distribution of standardized values for stolon emergence time observed in the epimutagenized population relative to control population.
Stolon emergency time density histogram of Z-test values from 284 hypomethylated lines.

A


Relative flowering time (day)
-10
-5
0
5
10
15
treatment
con
epi1
con1
epi2
con2
epi3
con3
epi4
con4
epi5
con5
Early flowering lines

B


Relative flowering time (day)
-10
0
10
20
treatment
con
epi1
con1
epi2
con2
Late flowering lines

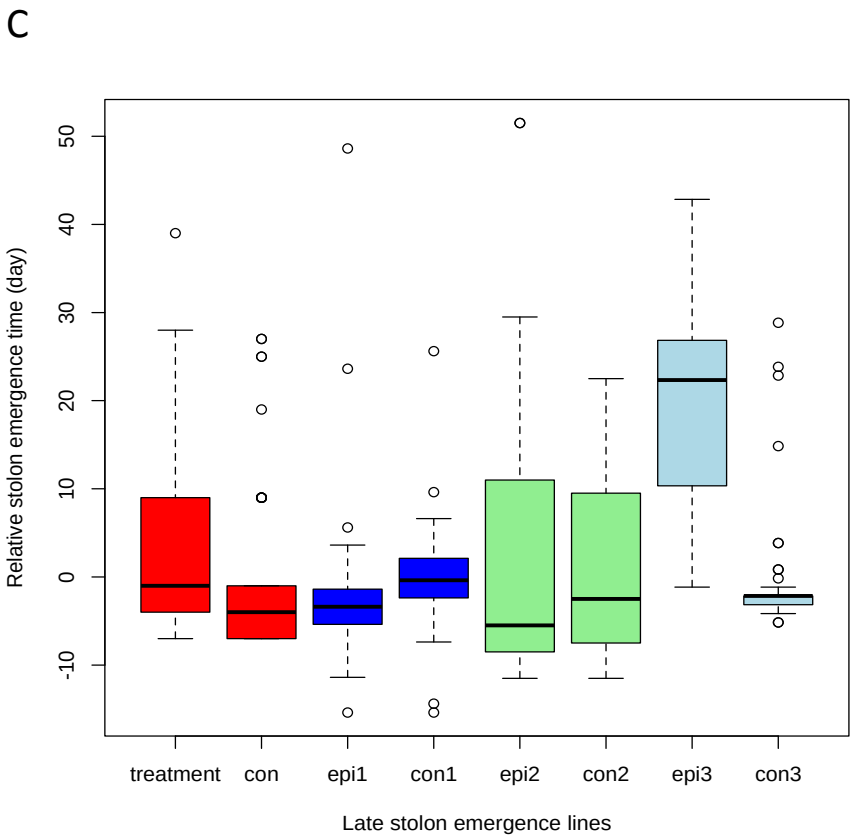


Figure 4.4 Quantitative traits distribution across generations.
The Y axis represents the relative value compared to the corresponding control (con) in early flowering time (A), late flowering time (B), and stolon emergence time (C) phenotypes. X axis, treatment and con are from the S0 generation epimutant lines and control lines; epi1 and con1 represent the S1 epimutant progenies and control lines; epi2 and con2 are S2 epimutant progenies and control lines; epi3 and con3 are S3 epimutant progenies and control lines; epi4 and con4 are S4 epimutant progenies and control lines; epi5 and con5 are S5 epimutant progenies and control lines.

normal distribution. In the S0 generation, the percentage of late stolon emergence individuals (21.55%) in the epimutant lines was twice the percentage of control lines (10%), similar to the flowering time trait in the S0 generation, no significant difference was detected (**Table 4.4**). Whereas in the S1 generation, this percentage was decreased to the level of about one third (3.6%) of control lines (10.9%) (**Figure 4.4 C**, **Table 4.4**). However, two extremely late stolon emergence lines (LS-54, LS-56) were found in the S1 generation (**Table 4.1**). These two individuals were selected and self-fertilized to generate the S2 generation.

These initial S0 and S1 results provide evidence of the effect of selection and the inheritance of flowering time through meiosis. The priority was given to the most extreme individuals in the family regardless of the mean difference compared to control. Therefore, two early flowering lines EF1, EF4 and two late flowering lines LF7 and LF8 were retained as they contained the earliest and latest flowering individuals. In total, three or four individuals with extremely early or late flowering time from each of these four families were selected and self-fertilized to produce the S2 generation (**Table 4.1**).

4.4.2 The transmission of flowering time and stolon emergence time in selected lines from S2 to S3 generations

In the seven S2 generation families identified with the early flowering trait, there were two families, EF4-9 and EF4-14, which had significantly earlier flowering time ($p < 0.05$) (**Table 4.1**). It was observed that 22.9% contained early flowering individuals, with only 13.6% of early flowering individuals in the control (**Table 4.2**). No significant difference was observed between the two combined early flowering families (EF1 and EF4) and control ($p = 0.28$). Since the early flowering trait was largely contributed by the families of EF4-9 and EF4-14, the five individuals with the most extreme early flowering phenotypes EF4-9-6, EF4-9-10, EF4-9-13, EF4-14-3, and EF4-14-7 were selected from these two families for study in future generations (**Table 4.1**).

By contrast, in the seven S2 generation families selected for the late flowering trait, all failed to transmit the late flowering phenotype (**Table 4.1**, **Figure 4.4 B**). There was no significant difference detected in the S2 generation between the late flowering and control families (**Table 4.3**). Thus, the late flowering phenotype was discontinued in the selection process.

In the two late stolon emergence families selected from the S2 generation, the family LS-56

Table 4.4 Composition of the late stolon emergence lines in each seed generation in epimutant and control lines.

Generation	Late stolon emergence epimutant			Control			$\chi 2$	p value
	Number of Late stolon emergence lines	Number of lines	%	Number of Late stolon emergence lines	Number of lines	%		
S0	61	283	21.6	5	50	10	3.57	0.06
S1	2	55	3.6	6	55	10.9	2.16	0.14
S2	15	60	25	8	32	25	0.00	1.00
S3	42	52	80.8	4	45	8.9	49.99	<0.001

Late stolon emergence lines were determined by the number of lines with stolon emergence time greater than the mean of the control plus one standard deviation (mean+SD).

had significantly late stolon emergence time whereas the family LS-54 had no significant difference in emergence time (**Table 4.1**). When both S2 generation families LS-54 and LS-56 were combined, the percentage of late stolon emergence individuals were equivalent to the control lines (25.0%, **Table 4.4**). Furthermore, no significant difference was observed between the late stolon emergence families and control (**Table 4.4**). Nevertheless, two extremely late stolon emergence individuals (LS-56-10, LS-56-25) from the family LS-56 were observed, and these two individuals were selected for the subsequent study of the transmission of late stolon emergence time to S3.

By the S3 generation, the families of all five selected early flowering lines (EF4-9-6, EF4-9-10, EF4-9-13, EF4-14-3, and EF4-14-7) had significantly earlier flowering time when compared to the control ($p < 0.05$) (**Table 4.1**). When combining all five families selected, the percentage of early flowering individuals in the early flowering epimutant lines (74.5%) was nearly six times the percentage of control lines (13.3%, **Table 4.2**) and a highly significant difference ($p < 0.001$) was detected between early flowering families and the control. Three of the earliest flowering individuals (EF4-9-13-15, EF4-14-7-6, EF4-14-7-14) were selected to evaluate transmission of this trait in one more generation. Similarly, the progenies of two selected late solon emergence families (LS-56-10, LS-56-25) also indicated significantly late stolon emergence time ($p < 0.05$) (**Table 4.1**). The percentage of late stolon individuals in the late stolon emergence epimutant lines (80.8%) was also highly significant ($p < 0.001$) at nine times the percentage of control lines (**Table 4.4**).

4.4.3 The transmission of early flowering time from S4 to S5 through meiosis and mitosis

The transmission of the early flowering time phenotype was continued in the subsequent S4 and S5 generations. Consistently, all S4 generation families derived from the three early flowering lines significantly flowered earlier compared to control lines ($p < 0.05$) (**Table 4.1**). The percentage of early flowering individuals in the early flowering epimutant lines (54.4%) was nearly three times the percentage of control lines (20.0%, **Table 4.2**) and a significant difference was detected between these three early flowering families and control lines ($p = 0.01$).

In order to further test the stability of the early flowering trait transmission in selected lines through both meiosis and mitosis, three extremely early flowering individuals EF4-9-13-15-18, EF4-14-7-6-12, and EF4-14-7-14-12 from the S4 families EF4-9-13-15, EF4-14-7-6, and

EF4-14-7-14 were used to produce the fifth generation S5 families through seed reproduction. Mitotic heritability was evaluated in clonal daughter plant families EF4-9-13-15', EF4-14-7-6', EF4-14-7-14', and control' generated from each of the S4 mother plants EF4-9-13-15, EF4-14-7-6, EF4-14-7-14. A highly significant difference was detected between the three early flowering families and control ($p < 0.001$) in the S5 generation, with 50.0% of the lines in these three families expressing the early flowering trait (**Table 4.2**). All three families (through both meiosis and mitosis) had significantly earlier flowering time compared to control ($p < 0.05$) (**Table 4.5**). In the clonally propagated generation through mitosis, flowering time in early flowering families was around six to ten days earlier than in control lines. These vegetatively propagated early flowering plants were on average advanced 5 days earlier than the lines which were propagated through seed.

4.4.4 Germination time in early flowering and control lines

A germination test was conducted in early flowering and control lines in order to explore if the early flowering lines were based on advanced plant development due to earlier germination. In total, five generations of early flowering lines derived from the EF4 line and eight control lines were evaluated for germination time, with each line consisting of 6 to 20 seeds. The average day to germination (DTG) of early flowering lines in 98 individuals was 12.6 with SD 2.6, which was similar to the average and SD of DTG (13.3 and 3.2 respectively) observed in the control lines (**Table 4.6**). No significant difference was found ($p = 0.151$) between the germination time of the early flowering and control lines in these two groups. Thus, the early flowering trait does not appear to be caused by advanced germination.

4.5 Discussion

In this study, an induced hypomethylated epi-mutant population of genetically identical individuals (Chapter 3), through repeated selection of extreme phenotypic traits (such as early flowering) was shown to be heritable through meiosis. The epialleles induced by both genetic elements and environmental factors causing heritable traits is generally more difficult to characterize compared to alleles caused by DNA mutations (Paszkowski and Grossniklaus, 2011; Becker and Weigel, 2012). By growing the genetically identical plants under stable greenhouse environmental conditions and including controls in each generation, we aimed to address some of these issues. The generation of a hypomethylated population of *F. vesca* and the demonstration that alterations in DNA methylation patterns can expand phenotypic

Table 4.5 Flowering time in the S5 generation in early flowering lines through meiosis and mitosis.

Meiosis				Mitosis			
Families	Size	Mean	SD	Families	Size	Mean	SD
EF4-9-13-15-18	12	75.9*	2.6	EF4-9-13-15’	15	17.3***	4.0
EF4-14-7-6-12	12	76.2*	2.4	EF4-14-7-6’	15	20.9*	4.4
EF4-14-7-14-12	12	74.3***	2.3	EF4-14-7-14’	15	21.3*	5.2
Control	24	78.5	2.8	Control’	15	27.1	7.1

*p <0.05 *** p < 0.001 “Size” = the number of individuals.

Table 4.6 Germination time in early flowering and control lines.

	Generation	Families	Mean (day)	SD	Size
Early	S1	EF4	11.9	1.6	20
	S2	EF4-9	12.3	2.8	15
	S3	EF4-9-13	11.8	3.8	13
	S3	EF4-14-7	11.9	1.9	16
	S4	EF4-14-7-14	11.3	1.9	16
	S5	EF4-9-13-15-18	15.6	3.0	12
	S5	EF4-14-7-14-12	16.8	3.4	6
Control		Control 1	14.9	1.7	9
		Control 2	14.7	4.1	9
		Control 3	12.1	3.2	19
		Control 4	10.1	1.7	18
		Control 5	13.8	5.0	10
		Control 6	14.3	2.9	14
		Control 7	15.5	3.8	14
		Control 8	13.4	3.4	19

“Size” = the number of individuals. P < 0.05.

variance of quantitative characters in genetically uniform individuals provides a resource to test if DNA methylation polymorphisms can behave as epialleles that are heritable and subject to selection.

That selection of a phenotype based on epialleles is not always stable is generally known and has been reported (Hirsch et al., 2012). However, the plasticity of phenotypic traits has also been shown to be subject to selection (Scheiner, 1993; Van Kleunen and Fischer, 2003; Callahan and Pigliucci, 2005) and has not been widely investigated in horticulture crops. Therefore, the selection of epigenetic variants was conducted in flowering time and stolon emergence time in *Fragaria vesca* in our research. In our study, 5-azaC was used as an epimutagen to induce the S0 hypomethylated population. This tool enables breeders to potentially increase selectable and heritable phenotypic variants of interest. In our selection from the S1 generations, we considered the extreme individuals rather than the families, which is different from the selection method used in classical breeding methods. In another study, selection was also made not only based on the mean but considered the variation and the extreme lines (Zhang et al., 2013). Although the initial production of the epimutants differed, the subsequent selection method of variants in our study was similar to the artificial selection for respiration and energy use efficiency reported in canola (*Brassica napus*) (Hauben et al., 2009). The starting material in that study was selected based on plants with the highest and lowest respiration rate. Four lower respiration lines and three higher respiration lines were obtained after four rounds of selection of high and low respiration lines. This strategy was used in our work because the segregation of epialleles in the variants provided extremely early or late flowering time lines, or late stolon emergence lines without significant changes in the means due to the offset of two extremes compared to the control population. Therefore, only considering the family means might run the risk of omitting these variant phenotypes.

However, in the following S2 generation selection, only the families with statistically distinct phenotypic traits were considered and extreme individuals from these potential families were selected to continue the propagation through meiosis. This is because at this stage, the related epialleles across the genome is expected to be more stable and the individuals with potential epialleles are expected to show consistent phenotypic traits. All the early flowering and late stolon lines selected in the S2 generations revealed stable transmission of epialleles across one generation (stolon, LS-56) or two consecutive generations (early flowering, EF4-9 and

EF4-14). However, the selected seven latest flowering individuals from the S1 generation failed to transmit to the next generation. This suggest the epialleles related to late flowering were reset through meiosis to cause a transient phenotypic change, as research indicated reversion of DNA methylation patterns affect the ability of epialleles to be subject to selection (Becker et al., 2011; Hirsch et al., 2012). Alternatively, no epialleles were present and the late flowering phenotype was induced by the toxic effect of 5-azaC or stress which disturbed the flowering time in the S0 generation (Jordan et al., 2015). In the S3, S4 and S5 generations, the early flowering trait was stably inherited. Thus, the early flowering related heritable epigenetic components may be stabilized under several rounds of phenotypic selection.

Several previous studies reported the inheritance of flowering time using epiRILs (Johannes et al., 2009; Roux et al., 2011; Zhang et al., 2013). This phenotypic inheritance of flowering time was correlated with the underlying DNA methylation alterations (Johannes et al., 2009). In our study, the early flowering line EF4, derived from 5-azaC treated genetically identical seeds, possessed the lowest methylation level detected by MSAP among all of the 22 epimutants selected to study, and the sequencing result verified no SNP observed in genes controlling flowering time in this line (Chapter 3). To the best of our knowledge, there have been no studies of this nature in *Fragaria* or other members of the Rosaceae family. In a canola energy use efficiency study, the AFLP test did not suggest any genetic polymorphism while MSAP revealed global hypomethylation and well-maintained differentially methylated fragments in selected lines (Hauben et al., 2009).

Through the germination time test, we verified a similar germination time between early flowering lines and control lines. This indicated the early flowering phenotype was not related to advanced development through earlier germination of the early flowering lines. Flowering is a multigene controlled complex process in plants, and the flowering pathway has been reported in numerous studies at the gene and molecular level in *Arabidopsis* (Simon et al., 1996; Samach et al., 2000; Yoo et al., 2005; Andrés and Coupland, 2012) and *Fragaria* (Sønsteby and Heide, 2007, 2008; Mouhu et al., 2009; Koskela et al., 2012). In *F. vesca* Hawaii 4, early flowering occurred under long day photoperiod by upregulating the *FvFT1*, which further promoted the regulation of *FvAP1* and *FvFUL1* for early flowering (Koskela et al., 2012). In Chapter 6, these genes were evaluated in our early flowering lines to determine the molecular mechanism of this phenotype.

Early flowering time is an important characteristic in strawberry breeding and production. From an evolutionary standpoint, flowering time genes are often targets of selection which can be fixed to increase fitness in order to better adapt to the environment (Roux et al., 2006). For the first time, our research verified the inheritance of early flowering through both meiosis and mitosis in strawberry. Most horticultural crops including strawberry are commercially propagated clonally. Therefore, the verification of flowering time transmission to mitotic daughters will make it possible for strawberry plants to be produced from epigenetic variants possessing the early flowering trait. Using the epigenetic approach independent of DNA primary sequence changes is a new method to improve complex traits in plant development and breeding.

4.6 Conclusions

We used the Rosaceae family model plant *F. vesca* epimutant lines of early flowering, late flowering, and late stolon emergence to investigate the transmission of phenotypic traits based on selection of induced epigenetic variation. After two rounds of selection of early flowering lines and late stolon emergence lines, the data indicated these two traits were successfully transmitted to subsequent generations. Inheritance of the early flowering trait was further examined over several generations (S3-S5) to develop stable populations with an altered mean flowering time. The transmission of early flowering time was demonstrated through both meiosis and mitosis. However, the transmission of the late flowering phenotype was not observed in the S2 generation. Here the variants possessing early flowering phenotypes initially identified from a 5-azaC induced hypomethylated population of *F. vesca* were successfully inherited through five generations. This opens the door to a new approach of utilizing epigenetic variation as a breeding tool in horticulture crops for inheritance through either generative or vegetative propagation methods. The DNA methylation profiles underlying the stable early flowering lines were further studied in Chapter 5.

5.0 ASSESSMENT OF THE METHYLATION PATTERN IN WOODLAND STRAWBERRY USING ESTABLISHED AND NOVEL MSAP APPROACHES

5.1 Abstract

The repeated selection of an induced hypomethylated epimutant population for early flowering time demonstrated this trait was heritable over five successive generations, resulting in the development of lines with distinct flowering habits. This suggests potential epialleles might exist in these early flowering lines with heritable flowering time. In order to explore the DNA methylation patterns underlying this trait that are unrestricted to the 5'-CCGG-3' context (the standard MSAP detected using isoschizomers *Hpa* II / *Msp* I), three pairs of new isoschizomers were applied to the MSAP method to detect DNA methylation levels and localization to their respective restriction sites. The isoschizomers *Tfi* I / *Pfe* I successfully detected DNA methylation at CG, CHG, CHH sites. This newly modified MSAP method was further used to explore DNA methylation patterns in two early flowering lines, EF lines 1 (P2) and EF lines 2 (P3) and control lines (P1). A significant difference was detected between P1 and P2 using the novel MSAP method that exploited the *Tfi* I / *Pfe* I combination, while the standard MSAP method found a significant difference between P1 and P3. Using the standard MSAP, the symmetric CG and CHG methylation was maintained over generations in the early flowering lines based on the clustering in P2 and P3, whereas using the novel MSAP approach, the asymmetric CHH methylation pattern was not maintained over generations indicated as less clustering of P2 and P3. This maintenance in CG and CHG methylation over generations suggests the early flowering phenotype might be related to DNA methylation alterations induced by 5-azacytidine at the CG or CHG sites. Combining these two techniques provides greater ability to find the underlying epialleles controlling the identified early flowering trait.

5.2 Introduction

In addition to methylation occurring at the symmetrical methylated sites CG and CHG (H = A, C, T) in DNA, non-symmetrical cytosine (CHH) was also observed using bisulfite sequencing in plants (Meyer et al., 1994). The mechanism for symmetrical DNA methylation is maintenance of DNA methylation. The newly unmethylated synthesized strand after DNA replication is methylated based on the methylation status of the CG sites on the leading strand by DNA methyltransferase MET1, recruited by VIM1, a methylcytosine-binding protein.

CHG methylation is maintained by the DNA methyltransferase CMT3 with the involvement of H3K9me2 methyltransferase SUVH4 / KYP (Teixeira and Colot, 2010). The asymmetric methylated sites commonly exist in transposons and repetitive elements (Chen et al., 2010; González et al., 2011). The mechanism of this asymmetrical methylation appears to be based on RNA-directed DNA methylation (RdDM) in plants through DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) and RNA polymerases catalyzing the small RNA base paired DNA, as first explained in tobacco plants (Wassenegger et al., 1994). Recently reports in *Arabidopsis thaliana* and *Brassica oleracea* indicated lower DNA methylation levels in CHH compared to the CG and CHG methylation sites (Cokus et al., 2008; Parkin et al., 2014). In *Arabidopsis* the methylation levels of CG, CHG and CHH was 24%, 6.7% and 1.7%, respectively.

There are several different techniques to detect DNA methylation on a large scale. Methylated DNA immunoprecipitation (MeDIP) using antibodies to identify methyl-cytosines can be coupled with high-resolution microarrays to detect methylated sequences, especially at CG rich sequences (Weber, 2005). The first high resolution methylome in plants was reported using this method in *Arabidopsis* (Zhang et al., 2006). The bisulfite sequencing method combining bisulfite treatment and next generation sequencing can achieve accurate methylation profiling (Cokus et al., 2008). It is known as the gold standard method with single base resolution to measure whole genome DNA methylation. However, these methods are not practical for DNA methylation screening in a population due to the need for high-throughput followed by complicated bioinformatics analysis. By contrast, using the methylation sensitive restriction enzymes to measure DNA methylation polymorphism in a large number of lines is cost-effective and efficient.

Isoschizomers are enzymes having the same recognition site except that one enzyme is methylation sensitive and the other is methylation insensitive (Bird and Southern, 1978). In the standard MSAP method, the isoschizomers *Hp*a II and *Msp* I were often exploited as frequent cutters to replace *Mse* I in AFLP but detect genome-wide DNA methylation (Cedar et al., 1979; Vos et al., 1995; Reyna-López et al., 1997). Due to the efficiency and low cost, numerous studies used this method to determine cytosine methylation in different fields from tissue culture (Chakrabarty et al., 2003; Xu et al., 2004), environmental stress (Shan et al., 2013; Rico et al., 2014), to ecology (Bossdorf et al., 2010; Schulz et al., 2013), and 5-azacytidine (5-azaC) treated plant material (Sano et al., 1990; Fieldes, 1994b; Marfil et al.,

2012). Since the recognition site of *Hp*a II and *Msp* I is 5'-CCGG-3', this method can only investigate DNA methylation based on 5'-CCGG-3'. Therefore, the output is limited to the cytosine methylations in the DNA contexts of CG and CHG (**Table 5.1**) without any knowledge of cytosine methylation on CHH. Besides the isoschizomers *Hp*a II and *Msp* I used in the MSAP method to study DNA methylation, isoschizomers *Acc65* I and *Kpn* I which recognize the 5'-GGTACC-3' site was used in screening of tissue-culture induced genetic and / or epigenetic variation in barley (Bednarek et al., 2007). However, studies attempting to use other isoschizomers and characterize DNA methylation on CHH DNA sequences have not been reported.

In a previous study (Chapter 3), we detected the DNA methylation profiles at 5'-CCGG-3' sites with the standard MSAP method in a hypomethylated population using isoschizomers *Hp*a II and *Msp* I together with *Eco*R I. In order to screen for additional cytosine methylation at CHH at the genome wide level, three new isoschizomers *CviA* II / *Nla* III, *Tfi* I / *Pfe* I, and *Dpn* II / *Sau3A* I were tested as frequent cutters in this research. The test indicated clear, scorable fragments generated with the isoschizomers *Tfi* I / *Pfe* I. Therefore, the isoschizomers *Tfi* I / *Pfe* I together with *Hp*a II / *Msp* I were further used to study the DNA methylation profiles at all CG, CHG, and CHH methylation context at specific restriction sites in a subset of early flowering lines generated in Chapter 4. Using a suite of isoschizomers provides a more comprehensive tool to identify the epialleles controlling phenotypic traits of interest.

5.3 Materials and Methods

5.3.1 Plant materials

Twenty-eight epimutant lines from a hypomethylated population of the eighth generation of *F. vesca* ssp. *vesca* accession Hawaii 4 (H4S8) which was generated from 5-azaC treatment (Chapter 3) were used to test the three pairs of isoschizomers (**Table 5.1** *CviA* / *Nla* III, *Tfi* I / *Pfe* I, and *Dpn* II / *Sau3A* I) using the MSAP method. The successfully modified MSAP method was further applied to control H4S8 lines and two early flowering lines EF lines 1 and EF lines2 (**Figure 5.1**). These were collected from different generations (generation 1, generation 2, and generation 3) of progeny from one early flowering line EF4 (originally named ERFv153) in the hypomethylated population treated with 5-azaC. EF4 did not possess genetic polymorphism to its control population in assays for SNP using short read based

Table 5.1 Isoschizomers recognition sites and methylation sensitivities.

	Isoschizomers	Recognition sites	Sensitivity to methylation
First pair	*Hpa* II	5-C/CGG-3 3-GGC/C-5	CG CHG
	Msp I	5-C/CGG-3 3-GGC/C-5	CG CHG
Second pair	*CviA* II	5-C/ATG-3 3-GTA/C-5	CHH
	Nla III	5-CATG/-3 3-/GTAC-5	CHH
Third pair	*Tfi* I	5-G/AWTC-3 3-CTWA/G-5	CG CHG CHH
	Pfe I	5-G/AWTC-3 3-CTWA/G-5	CG CHG CHH
Forth pair	*Dpn* II	5-/GATC-3 3-CTAG/-5	CG CHG CHH
	Sau3A I	5-/GATC-3 3-CTAG/-5	CG CHG CHH

Forward slashes indicate enzymes cutting sites
W means A or T

whole genome sequencing (Chapter 3).

5.3.2 Assessment of DNA methylation polymorphism using modified Methylation Sensitive Amplified Polymorphisms (MSAP)

The DNA extraction procedure was exactly the same as described in Chapter 3. Three new pairs of isoschizomers *CviA* / *Nla* III, *Tfi* I / *Pfe* I, and *Dpn* II / *Sau3A* I together with *Bgl* II were used to digest genomic DNA extracted from leaf material. The recognition sites and sensitivity to methylation context are listed in **Table 5.1**. The modified MSAP protocol was the same as the MSAP protocol described using *Hpa* II / *Msp* I with *Eco*R I except the enzymes and corresponding NEB buffer were replaced. The adaptors and primers used in these three modified MSAP are listed in **Table 5.2**. The amplified MSAP products were resolved using exactly the same method as the MSAP protocol previously described in Chapter 3.

5.3.3 Statistical analysis

MSAP profile analysis was determined using Principal Coordinate Analysis (PCoA) and Analysis of Molecular Variance (AMOVA) as described in Chapter 3. Loci with at least 5% (the proportion of type II, III bands) methylation levels were defined as methylation-susceptible loci. Polymorphic methylation-susceptible loci were defined when at least two individuals were non-methylated.

5.4 Results

5.4.1 DNA methylation profiles using three new pairs of isoschizomers

Three new isoschizomers, namely *CviA* / *Nla* III, *Tfi* I / *Pfe* I, and *Dpn* II / *Sau3A* I, together with *Bgl* II were used in the MSAP assay in 28 genetically uniform epimutant lines that had been treated with 5-azaC. The MSAP banding patterns revealed very informative polymorphic patterns in the isoschizomeric combination of *Bgl* II / *CviA* II and *Bgl* II / *Nla* III (**Figure 5.2 A**). A total of 72 MSAP loci were amplified with only one primer pair combination. Although the recognition site of isoschizomers *CviA* II / *Nla* III was the same (5'-CATG-3'), the cleavage site of *CviA* II occurred after cytosine while the cleavage site of *Nla* III occurred after guanine. As a result of this, it is difficult to identify the fragments corresponding to the same loci in these two combinations *Bgl* II / *CviA* II and *Bgl* II / *Nla* III, potentially interfereing with the scoring of amplified bands.

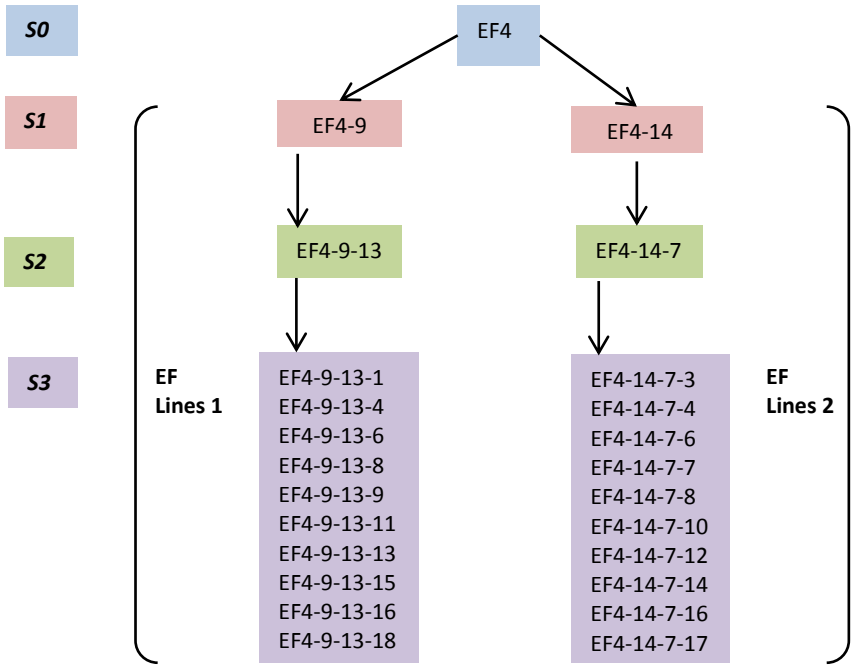


Figure 5.1 The overview of two early flowering lines (EF lines 1, EF lines 2) composed of four generations.
EF4 is an early flowering line selected from a hypomethylated population treated with 5-azaC.

Table 5.2 Sequences of adaptors and primers used for pre-selective amplification and selective amplification in MSAP using three pairs of isoschizomeric combinations. *Bgl* II / *CviA* II and *Bgl* II / *Nla* III, *Bgl* II / *Tfi* I and *Bgl* II / *Pfe* I, *Bgl* II / *Dpn* II and *Bgl* II / *Sau3A* I were the three combinations.

Adapters/primers	*Bgl* II	*CviA* II
Adapter	5-CTCGTAGACTGCGTACC-3	5-GACGATGAGTCTAGAA-3
	3-CATCTGACGCATGGCTAG-5	3-CTACTCAGATCTTTA-5
Pre-amplification primers	5-GACTGCGTACCGATCT A-3	5-GATGAGTCTAGAAATG C-3
Selective primers	5-GACTGCGTACCGATCT AGT-3	5-GATGAGTCTAGAAATG CAT-3
Adapters/primers	Bgl II	*Nla* III
Adapter	5-CTCGTAGACTGCGTACC-3	5-TTGACGATGAGTCTAGAACATG-3
	3-CATCTGACGCATGGCTAG-5	3-TGCTACTCAGATCTT-5
Pre-amplification primers	5-GACTGCGTACCGATCT A-3	5-GATGAGTCTAGAACATG C-3
Selective primers	5-GACTGCGTACCGATCT AGT-3	5-GATGAGTCTAGAACATG CAT-3
Adopters/primers	*Bgl* II	*Tfi* I/*Pfe* I
Adapter	5-CTCGTAGACTGCGTACC-3	5-GACGATGAGTCTAGAA-3
	3-CATCTGACGCATGGCTAG-5	3-CTACTCAGATCTTTWA-5
Pre-amplification primers	5-GACTGCGTACCGATCT A-3	5-GATGAGTCTAGAAAWTC C-3
Selective primers	5-GACTGCGTACCGATCT AGT-3	5-GATGAGTCTAGAAAWTC CAT-3
Adopters/primers	*Bgl* II	*Dpn* II/*Sau3A* I
Adapter	5-CTCGTAGACTGCGTACC-3	5-GACGATGAGTCTAGAA-3
	3-CATCTGACGCATGGCTAG-5	3-CTACTCAGATCTTCTAG-5
Pre-amplification primers	5-GACTGCGTACCGATCT A-3	3-GATGAGTCTAGAAGATC T-5
Selective primers	5-GACTGCGTACCGATCT AGT-3	3-GATGAGTCTAGAAGATC TAC-5

The second isoschizomeric combination *Bgl* II / *Tfi* I and *Bgl* II / *Pfe* I was assayed in the same sample. According to the report in the restriction enzyme database REBASE (http://rebase.neb.com/cgi-bin/msrecget) (Roberts et al., 2005), enzyme *Tfi* I is sensitive to methylation while its isoschizomer *Pfe* I is nonsensitive to methylation. There are a total of three types of bands amplified using the isoschizomers *Tfi* I / *Pfe* I in the MSAP assay. When no cytosine methylation of the recognition sequence 5'-GAWTC-3' (nonmethylation) occurred, both isoschizomers *Tfi* I / *Pfe* I are able to cut this sequence and generate the type I band which is scored as 1/1. When cytosine is methylated on one of the strands of the recognition sequence 5'-GAWTC-3' (hemimethylation), the enzyme *Tfi* I can cut this sequence but its isoschizomers *Pfe* I cannot cut. This generates the type II band and is scored as 1/0. When both cytosines are methylated on the two sequence strands (full methylation) this is referred to as type III band. As expected, all three types of bands type I, type II, and type III were clearly amplified and resolved as shown in **Figure 5.2 B**, with a total number of 57 MSAP loci amplified using one primer pair combination.

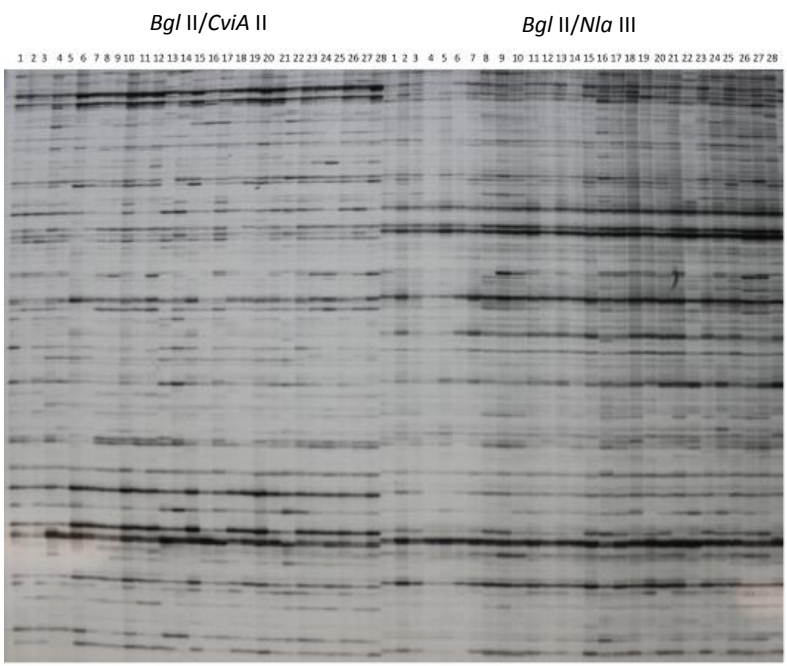
The last isoschizomeric combinations assayed were *Bgl* II / *Dpn* II and *Bgl* II / *Sau3A* I. The recognition sequence of these isoschizomers is 5'-GATC-3'. Theoretically, similar to the second isoschizomeric combination, there are three types of methylations corresponding to three types of bands. However, there were few bands amplified (**Figure 5.2 C**). The REBASE database (http://rebase.neb.com/cgi-bin/msrecget) (Roberts et al., 2005) indicated the *Dpn* II enzyme can cut the recognition sequence while its isoschizomer *Sau3A* I cannot cut when in a hemimethylated or fully methylated state.

Within these three new isoschizomers, only combination *Bgl* II / *Tfi* I and *Bgl* II / *Pfe* I successfully generated scorable polymorphic DNA methylation bands. This novel approach (named novel MSAP) was further applied to two early flowering lines EF lines 1 (12 individuals) and EF lines 2 (12 individuals) (**Figure 5.1**), and H4S8 control lines (10 individuals).

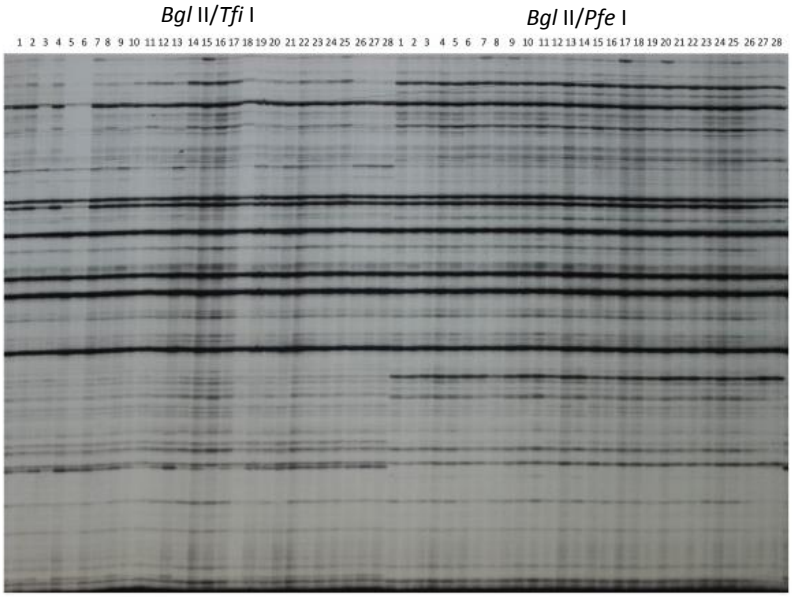
5.4.2 DNA methylation profiles in early flowering lines and control lines using standard MSAP and novel MSAP

A total of 166 MSAP loci were amplified and scored using six primer pairs with isoschizomer combinations *Bgl* II / *Tfi* I and *Bgl* II / *Pfe* I (**Table 5.3**). In order to compare with the result of the standard MSAP, four primer pair combinations (**Table 5.4**) were used in isoschizomeric

A



B



C

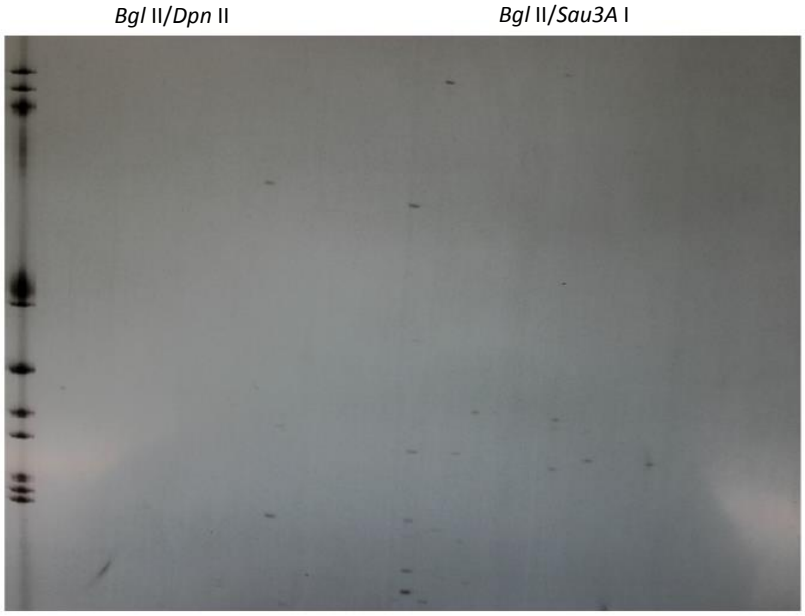


Figure 5.2 MSAP profiles generated using three new pairs of isoschizomeric combinations. 1-28: 28 epimutant lines used to test. A: *Bgl* II / *CviA* II and *Bgl* II / *Nla* III, B: *Bgl* II / *Tfi* I and *Bgl* II / *Pfe* I, C: *Bgl* II / *Dpn* II and *Bgl* II / *Sau3A* I. The primer combinations used are shown in Table 5.2

combinations *Eco*R I / *Hpa* II and *Eco*R I / *Msp* I in the same EF lines 1, EF lines 2 and control H4S8 lines. A total of 142 loci were amplified and scored in this standard MSAP approach. In total, three types of bands were grouped together based on scoring from these two MSAP approaches. There are two cytosines in double-stranded *Tfi* I / *Pfe* I recognition site GAWTC, resulting in three types of bands: type I representing nonmethylation, type II representing hemimethylation, and type III representing full methylation. By contrast, there are more cytosine methylation patterns of double-stranded *Hpa* II / *Msp* I recognition site 5'-CCGG-3' as demonstrated in Chapter 3 in standard MASP. For the purpose of comparison, we combined both type III and type IV bands in *Hpa* II / *Msp* I as type III bands which represent both stranded cytosines methylation. Among the three groups studied (**Table 5.5**), the novel MSAP generated a higher percentage of nonmethylated bands (around 80%) compared to the standard MSAP (nearly 60%). This can be explained by the different cytosine numbers in the two isoschizomer recognition sequences. There are two cytosines; one is in a CG methylation context in isoschizomers *Hpa* II / *Msp* I recognition site 5'-CCGG-3', while *Tfi* I / *Pfe* I recognition site 5'-GAWTC-3' only has one cytosine (**Table 5.1**). In addition, in both novel and standard MSAP assays across all three groups, type I bands had the highest level, followed by type III bands, with type II bands at the lowest level.

Epigenetic differentiation among control line H4S8, EF lines 1 and EF lines 2 based on all the loci was distinguished according to Principal Coordinate Analysis (PCoA). In the standard MSAP, a total of 142 loci were partitioned into 53 methylation-susceptible loci (MSL) (37%) and 89 non methylated loci (63%). Only the 10 (19% of all MSL) polymorphic methylation-susceptible loci were used in the PCoA. By contrast, in the novel MSAP, a total of 166 loci were separated into 29 methylation-susceptible loci (17%) and 137 loci without methylation (83%). The number of polymorphic methylation-susceptible loci was 8 (28% of all MSL) and these loci were used in the PCoA. In the PCoA (**Figure 5.3 A**) of H4S8 lines (P1), EF lines 1 (P2) and EF lines 2 (P3) using the standard MSAP, the first two coordinates explained 63.8% of the variance. The P1 represented control H4S8 lines and was evenly spread out along the two coordinates. This indicated a similar pattern as reported in Chapter 3 control lines (**Figures 3.7**, **Figure 3.10)**. By contrast, most of the EF lines 1 (P2), EF lines 2 (P3) were distributed along the first coordinate. The significant difference in epigenetic variation among H4S8 lines (P1), EF lines 1 (P2), and EF lines 2 (P3) was detected (AMOVA, P value = 0.0035). Pairwise PhiST comparisons indicated H4S8 lines (P1) were significantly different from EF lines 2 (P3) (p value = 0.0166). Similar to the distribution of H4S8 lines (P1) of the

Table 5.3 Sequences of adaptors and primers used for pre-selective amplification and selective amplification in new MSAP method using *Bgl* II and isoschizomers *Tfi* I / *Pfe* I.

Adapters/primers	*Bgl* II	*Tfi* I / *Pfe* I
Adapter	5-CTCGTAGACTGCGTACC-3	5-GACGATGAGTCTAGAA-3
	3-CATCTGACGCATGGCTAG-5	3-CTACTCAGATCTTTWA-5
Pre-amplification primers	5-GACTGCGTACCGATCT A-3	5-GATGAGTCTAGAAAWTC C-3
Selective primers	5-GACTGCGTACCGATCT AGT-3	5-GATGAGTCTAGAAAWTC CAT-3
	5-GACTGCGTACCAATTC AGT-3	5-GATGAGTCTAGAAAWTC CTT-3
	5-GACTGCGTACCAATTC AGT-3	5-GATGAGTCTAGAAAWTC CCT-3
	5-GACTGCGTACCAATTC AGA-3	5-GATGAGTCTAGAAAWTC CAT-3
	5-GACTGCGTACCAATTC AGA-3	5-GATGAGTCTAGAAAWTC CTT-3
	5-GACTGCGTACCAATTC AGA-3	5-GATGAGTCTAGAAAWTC CCT-3

Table 5.4 Sequences of adaptors and primers used for pre-selective amplification and selective amplification in standard MSAP method using *Eco*R I and isoschizomers *Hpa* II / *Msp* I.

Adapters/Primers	*Eco*R I	*Hpa* II / *Msp* I
Adapter	5-CTCGTAGACTGCGTACC-3	5-GACGATGAGTCTAGAA-3
	3-CATCTGACGCATGGTTAA-5	3-CTACTCAGATCTTGC-5
Pre-amplification primers	5-GACTGCGTACCAATTC A-3	5-GATGAGTCTAGAACGG T-3
Selective primers	5-GACTGCGTACCAATTC ACT-3	5-GATGAGTCTAGAACGG TAA-3
	5-GACTGCGTACCAATTC ACT-3	5-GATGAGTCTAGAACGG TTG-3
	5-GACTGCGTACCAATTC ACA-3	5-GATGAGTCTAGAACGG TTA-3
	5-GACTGCGTACCAATTC ACA-3	5-GATGAGTCTAGAACGG TTT-3

Table 5.5 The summary of three types of bands in control lines and two early flowering lines (EF lines 1 and EF lines2).

	H4S8		EF lines 1		EF lines 2	
	Hpa II / *Msp* I	*Tfi* I / *Pfe* I	*Hpa* II / *Msp* I	*Tfi* I / *Pfe* I	*Hpa* II / *Msp* I	*Tfi* I / *Pfe* I
type I	83 (58.5%)	132 (79.5%)	82 (57.7%)	133 (79.9%)	81 (57.4%)	134 (80.8%)
type II	15 (10.5%)	8 (4.8%)	14 (9.8%)	9 (5.8%)	14 (9.8%)	10 (5.9%)
type III	44 (31.0%)	26 (15.7%)	46 (32.5%)	24 (14.3%)	47 (32.8%)	22 (13.3%)
total	142	166	142	166	142	166

Parentheses data represent the percentage when compared to the total number of bands.

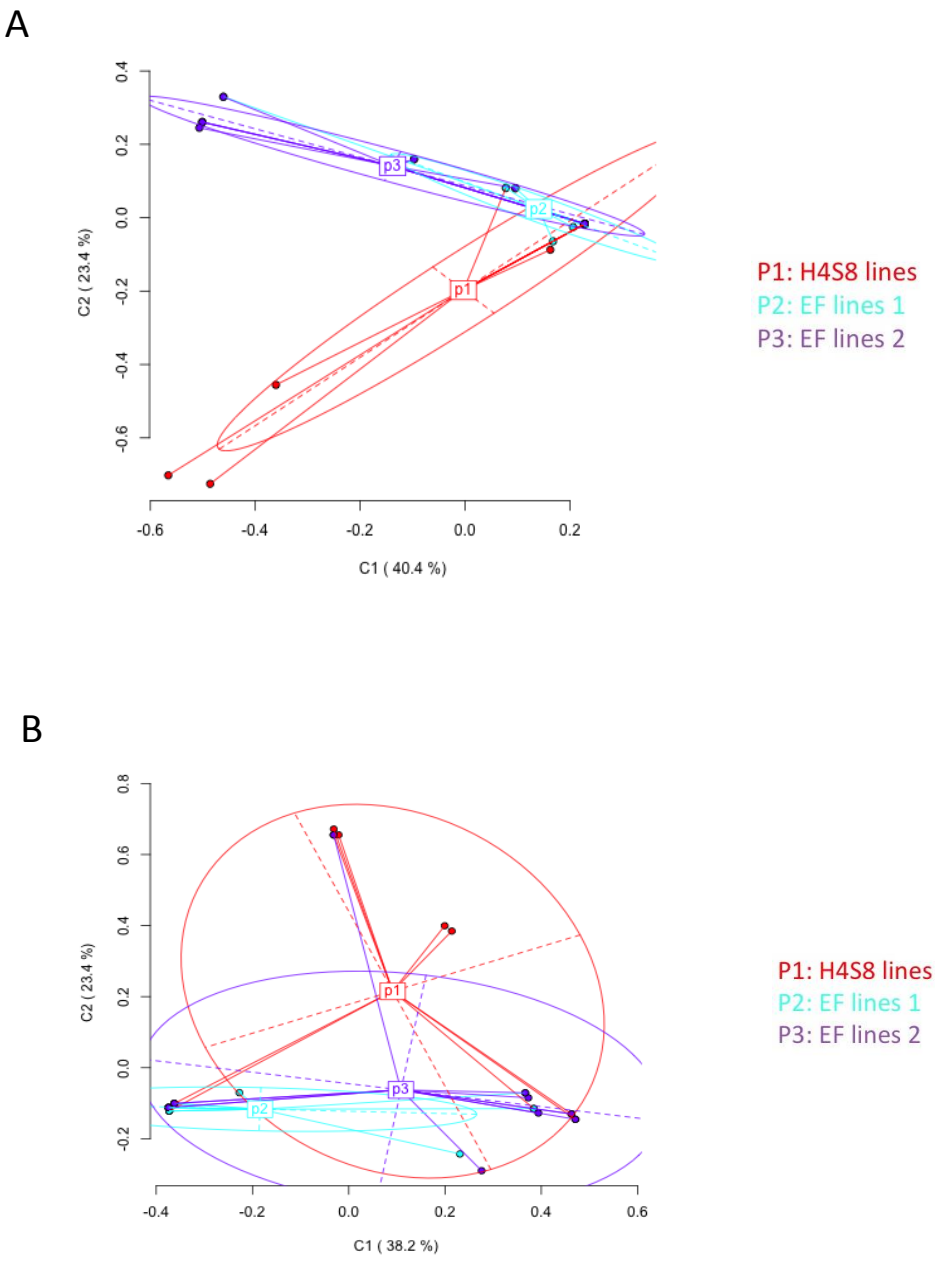


Figure 5.3 Principal Coordinates Analysis (PCoA) for DNA methylation differentiation between control lines and 5-azaC treatment lines using polymorphic methylation-susceptible loci (MSL) data.
The percentages in the first two coordinates (C1 and C2) are shown the amount of variance contributed by them (shown in brackets). Color-labeled P1, P2, P3 are the centroids of corresponding group. P1 represents control H4S8 lines, P2 represents early flowering lines 1, P3 represents early flowering lines 2. A: data obtained from standard MSAP B: data obtained from novel MSAP.

standard MSAP, in the novel MSAP, the H4S8 control was also evenly distributed along the first two coordinates that accounted for 61.6% of the variance, and most EF lines 1 (P2) and EF lines 2 (P3) were distributed along the first coordinate (**Figure 5.3 B**). A significant difference among the three groups was also observed (AMOVA, p value = 0.0125). Further Pairwise PhiST comparisons indicated a significant difference between H4S8 lines (P1) and EF lines 1 (P2) (p value = 0.0125). However, this contrasted with the standard MSAP experiment which detected significant differences between H4S8 lines (P1) and EF lines 2 (P3).

5.5 Discussion

Both MSAP and HPLC can be used to assess the DNA methylation levels across the genome with lower costs compared to bisulfite sequencing combined with next generation sequencing and Methylated DNA immunoprecipitation. However, while HPLC can detect the general overall level of methylated cytosine across the genome, no information of different patterns of cytosine contexts are provided (Johnston et al., 2005). By contrast, MSAP (standard) can determine the methylation at CG and CHG sites since the recognition sites are 5'-CCGG-3' (Reyna-López et al., 1997). As DNA methylation in plants occurs at three cytosine contexts: CG, CHG, and CHH, and different DNA methyltransferases that are involved in the *de novo* methylation and maintenance of DNA methylation of these contexts are related to particular biological functions (Teixeira and Colot, 2010; Kawashima and Berger, 2014), it is important to detect the DNA methylation profiles at both symmetric CG and CHG sites and the asymmetric CHH site. Therefore in this study, isoschizomers that can recognize the asymmetric CHH site were used in order to detect the cytosine methylation occurring at the asymmetric site, and further explore the maintenance of asymmetric cytosine methylation when the early flowering trait is passed to the following generations through meiosis. The isoschizomers *Tfi I* and *Pfe I* were found to cleave the genome DNA and generate clear scorable bands. Therefore, these isoschizomers were used to study the epigenetic variation in two early flowering lines. To our knowledge, this is the first attempt to reveal the polymorphisms of cytosine methylation based on novel isoschizomers *Tfi I* and *Pfe I*.

In maize, the majority of meiotically heritable methylated cytosines occur at symmetric CG methylation context (Lauria et al., 2014). Similarly, a study in *N. tabacum* found symmetric CG methylation was stably inherited for at least two generations while methylation at CHG, CHH was not maintained (Dalakouras et al., 2012). This appeared to be associated with

symmetric established cytosine methylation could be maintained through MET1 and CMT3 during each cycle of DNA replication. By contrast, methylation of CHH is not propagated based on template maintenance due to the asymmetrical sequence of CHH (Kankel et al., 2003; Henderson and Jacobsen, 2007). Thus, *de novo* methylation is required with the direction of small interfering RNA (Cao and Jacobsen, 2002). The material used in this study was derived from treating *F. vesca* seeds with the DNA demethylating reagent 5-azaC. The 5-azaC reagent is an analogue of cytidine resulting in hypomethylation without specific preference, targeting both symmetric and asymmetric cytosines (Jones et al., 1983). Therefore, after the 5-azaC demethylation of symmetric CG and CHG methylation, the demethylated status of cytosine can be maintained through cell division across generations. In contrast, if 5-azaC demethylates CHH asymmetric sites, this demethylation can be transmitted via cell reproduction but will be reverted into the original state (methylated CHH) through *de novo* methylation after removal of 5-azaC. Based on this, the clustering in P2 and P3 in standard MSAP might indicate symmetric CG, CHG methylation was maintained over generations. The reduced clustering of P2 and P3 according to the novel MSAP might indicate the asymmetric CHH methylation pattern was not maintained over generations. Based on the significant difference of DNA methylation variation among P1, P2, and P3 groups in standard and novel MSAP, it appears that DNA methylation in 5'-CCGG-3' (CG and CHG) accounted for the epigenetic variation in EF lines 2 (P3) while the DNA methylation in 5-GAWTC-3 (CG, CHG, and CHH) accounted for the epigenetic variation in EF lines 1 (P2).

To our knowledge, this research is the first attempt to assess DNA methylation using different isoschizomers including *CviA* / *Nla* III, *Tfi* I / *Pfe* I, and *Dpn* II / *Sau3A* I with *Tfi* I / *Pfe* I showing the best scorable patterns. The methylation level detected using isoschizomers *Tfi* I / *Pfe* I was less sensitive than *Hpa* II / *Msp* I of the standard MSAP. However, *Tfi* I and *Pfe* I are designed to detect more methylation patterns at the asymmetric CHH. The recognition site (5'-GAWTC-3') of *Tfi* I / *Pfe* I only contains one cytosine, the chance of CG is only 1/4 (25%) in 5'-GAWTC-3' and 9/16 (56%) of the possibility of CHH. However, in plants, methylation pattern in CG was higher while methylation in CHH is reduced. This may explain the reduced overall methylation level detected using the novel MSAP even though it can detect more CHH.

5.6 Conclusions

We assayed three pairs of new isoschizomers using the MSAP method to detect DNA methylation levels in specific sites. The isoschizomers *Tfi* I / *Pfe* I successfully detected DNA methylation at CG, CHG, CHH sites. This newly modified MSAP method was further used to explore DNA methylation patterns in two early flowering lines, EF lines 1 and EF lines 2, and control lines. Combined with the result of the standard MSAP method, the symmetric CG, CHG methylation was found to be maintained over generations in the early flowering lines, whereas the asymmetric CHH methylation pattern was not maintained. The methylation maintenance at CG and CHG indicated the heritable alterations in DNA methylation at these sites and

may be correlated with the early flowering phenotype. This provided the fundamental information for further detection of DNA methylation in early flowering lines based on single nucleotide resolution using whole genome bisulfite sequencing in Chapter 6.

6.0 CHARACTERIZING EARLY FLOWERING EPIGENETIC VARIANTS THROUGH RNA-SEQ AND SINGLE-BASE RESOLUTION METHYLOME IN STRAWBERRY

6.1 Abstract

Shoot Apical Meristems (SAM) collected from plants showing consistent inheritance of early flowering over five generations was characterized in this study. RNA-Seq analysis of the SAM from early flowering and control lines at three developmental stages indicated significant gene expression changes in stage 3. There were a total of five, 217, and 931 Differentially Expressed Genes (DEG) in stage 1, stage 2, and stage 3, respectively. Strawberry flowering time genes were identified through homologues search to *Arabidopsis* where 96 homologues found. None of the five DEG in stage 1 was related to flowering. By contrast, there were six genes in stage 2 and six genes in stage 3 involved in the flowering time pathway, with circadian clock gene *LATE ELONGATED HYPOCOTYL*/*CIRCADIAN CLOCK ASSOCIATED 1* (*LHY/CCA1*) observed in both stage 2 and stage 3. Most of these genes regulating flowering were in the photoperiod pathway. The up-regulation of downstream genes of flowering pathway *NUCLEAR FACTOR Y, SUBUNIT B2/HEME ACTIVATED PROTEIN 3B* (*NFYB2/HAP3B*), *FLOWERING LOCUS T* (*FT*), *FRUITFULL* (*FUL*) and *SEPALLATA3* (*SEP3*) likely contributed to floral transition in early flowering lines, especially *FT* with a six fold change. Since most of the DEG were detected in stage 3 and the genes *NFYB2/HAP3B, FT*, and *SEP3* were only differentially expressed in stage 3, the meristem of this stage was used to investigate the DNA methylation variation induced in early and control lines using whole genome bisulfite sequencing. Significantly different CG methylation levels in *FT* gene (early 32%, control 41%) and coding sequence regions (early 11%, control 5%) were observed but no differences in methylation were found between early and control lines when evaluated across all CG, CHG, CHH sites. This study examined molecular mechanisms of early flowering lines through gene expression and whole genome bisulfite sequencing. The single nucleotide based resolution of methylation levels of *F. vesca* at CG, CHG, and CHH sites was reported for the first time.

6.2 Introduction

Flowering is a multi-gene controlled process that is influenced by both internal (gibberellin, aging, autonomous pathways) and external (photoperiod and vernalization pathways) factors

(Greenup et al., 2009; Kim et al., 2009; Amasino, 2010). Plants need to pass through the juvenile stage to be competent to flower and can then respond to environmental factors that induce flowering (Poethig, 2003). Among the external factors, photoperiod plays a critical role in regulating flowering time and is associated with the circadian rhythms controlled by the circadian clock (Dunlap, 1999). Much of the regulatory information on flowering has been derived from *Arabidopsis*. *Arabidopsis* is a facultative Long Day (LD) plant, circadian clock genes *LHY / CCA1*, *TIMING OF CAB EXPRESSION1* (*TOC1*), *PSEUDO-RESPONSE REGULATOR7* / *9* (*PRR7* / *PRR9*) as well as downstream genes *GI*, *CO*, *FT* in the photoperiod pathway together regulate flowering (Fornara et al., 2010). *CO* is a key gene receiving the signal from the circadian clock and promoting *FT* transcription (Kardailsky et al., 1999). The expression of *CO* is repressed by *CYCLING DOF FACTORs* (*CDFs*) in leaves (Fornara et al., 2010). The signal *FT* protein moves from the leaves to the SAM. Together with *SUPPRESSER OF OVEREXPRESSION OF CO1* (*SOC1*) and *LEAFY* (*LFY*), they are the downstream integrator genes connecting multiple flowering pathways and the transcription levels of them affect flowering time (Lee and Lee, 2010).

The SAM produce the above-ground structure of plants as the sites of cell division and organ initiation. The fate of SAM changes from vegetative to inflorescence meristems when plant growth modes shift from the vegetative growth state to the reproductive state. The inflorescence meristems later initiate floral meristems (Mandel and Yanofsky, 1995), and the floral meristem identity genes *LFY*, *APETALA1* (*AP1*), *APETALA2* (*AP2*), *CAULIFLOWER* (*CAL*) determine flowering time (Levy and Dean, 1998). Besides being involved in carpel and fruit development, the *FUL* gene was reported to be upregulated during flowering transition at SAM and was found to be another meristem identity gene (Gu et al., 1998; Ferrándiz et al., 2000). The *SEP3* is a MADS transcription factor and was originally found as the floral organ identity gene involved in the development of petals, stamens, and carpels (Pelaz et al., 2000; Favaro et al., 2003). Studies also indicated *SEP3* expression was upregulated by *FT* and the role of *SEP3* in promoting flowering was by enhancing *AP1* (Pelaz et al., 2001; Hwan Lee et al., 2012).

In perennial plants such as strawberry, flowering time regulation is more complicated as plants experience repeated seasonal cycles of vegetative and reproductive stages that are controlled by seasonal flowering genes (Wang et al., 2009). There are seasonal flowering Short Day (SD) and perpetual Day Neutral (DN) genotypes in wild strawberry (*F. vesca*). The DN has been recognized as a temperature-dependent LD plants in some research (Sønsteby

and Heide, 2007; Stewart and Folta, 2010). Recent study found Hawaii 4 is a LD plant, but unlike the seed generation, the clonal daughters produced through stolon propagation were DN plants (Koskela et al., 2012). LD induced upregulation of *FvFT1*, as well as the floral meristem identity genes *FvAP1* and *FvFUL1* in Hawaii 4 were reported (Koskela et al., 2012). Flowering time as a polygenic controlled trait is regulated by many factors with more than 80 genetic loci involved in flowering (Ratcliffe and Riechmann, 2002). Diploid *F. vesca* has been recognized as a model system to study flowering time over the past few years (Mouhu et al., 2009; Mouhu et al., 2013). Gene expression analysis of a number of the 26 genes involved in each pathway were tested in everbearing and SD strawberry, with only a few genes showing differential expression between genotypes using real-time RT PCR (Mouhu et al., 2009). Floral development of *F. vesca* was comprehensively studied using next generation sequencing (Hollender et al., 2014). The availability of the *F. vesca* genome reference makes it possible to comprehensively study gene expression and the molecular basis of regulation of flowering time (Shulaev et al., 2011).

In addition to the genetic and environment regulation of flowering time, the epigenetic regulation of flowering time has been reported. For example, decreased DNA methylation induced early flowering in *Arabidopsis* (Burn et al., 1993), late flowering plants was also observed in a demethylated background of DNA methyltransferase (Soppe et al., 2000), and histone modification was related to *FLOWERING LOCUS C* (*FLC*) silencing during vernalization (Bastow et al., 2004; Sung and Amasino, 2004). *FLC* is a flowering repressor and is well studied in *Arabidopsis* vernalization and the autonomous pathway. In addition to histone modification, other chromatin modification was also found influencing flowering time through transcriptional regulation of *FLC* (He, 2012; Ietswaart et al., 2012). Since the MSAP method to assay DNA methylation is limited to 5'-CCGG-3' DNA sequence contexts, here we use Whole Genome Bisulfite Sequencing (WGBS) to determine the methylation status in different regions of the genome such as gene body, upstream, downstream. This methylation data were combined with RNA-Seq to determine the methylation profiles of DEG correlated with floral transition using early flowering lines obtained from five generations of selection in a hypomethylated *F. vesca* population. Based on our knowledge, this comprehensive transcriptome combined with a methylome study of the molecular basis of flowering time in *F. vesca* has not been previously reported. This information will provide a foundation for further study to find the epialleles controlling flowering time.

6.3 Materials and Methods

6.3.1 Plant materials and sampling

This study was based on two related early flowering lines EF4-9-13-15-18 (early1) and EF4-14-7-14-12 (early2) produced from the fifth generation of an early flowering line (EF4) generated using the DNA demethylation reagent 5-azacytidine treatment (Chapter 3, Chapter 4). Two wild type lines of the eighth generation Hawaii 4, H4S8-14 (control1) and H4S8-21 (control2) were used as control lines. The SAM from these four lines were dissected for RNA sequencing and methylome analysis.

A total number of 1200 seeds from each line were placed into petri dishes for germination. The growth condition was as previously described (Chapter 3), the greenhouse conditions were at 23 ± 2°C day and 18 ± 2°C night under an 18 / 6 h day/night photoperiod. The SAM were collected in three stages from each of these four lines (**Figure 6.1**). The stage 1 was defined when 2-3 leaves emerged, corresponding to 36 days after imbibing the seeds in petri dishes to induce germination. The stage 2 was defined when 5-6 leaves emerged, 50 days after seeding in petri dishes for germination. The stage 3 was defined when 9-10 leaves emerged, 64 days after seeding in petri dishes for germination. The SAM (1-2mm) were dissected from shoot tips with the aid of a stereoscope using tweezers and a dissecting blade. Since two biological replicates were applied at each stage, there were a total of 24 samples collected with 50 plants in each meristem sample analyzed using RNA-seq. A total of eight meristem samples for DNA methylation analysis were dissected from stage 3 only. All sampled meristems were placed in dry ice, put into liquid nitrogen before being transferred into -80°C. The meristems were stored at -80°C until RNA and DNA extraction. Twelve plants from each line were used to score the flowering time ensuring a consistent phenotype (Chapter 4).

6.3.2 RNA isolation, RNA-Seq library preparation and sequencing

Total RNA was isolated using the Plant/Fungi total RNA purification Kit (Norgen, Thorold, ON, Canada). RNA quantification was performed using Qubit 2.0 Fluorometer and the Qubit RNA BR Assay Kits (Invitrogen, Eugene, OR, USA) according to the manufacturer's instructions. The integrity of total RNA was assessed via the RNA 6000 Nano labchip on 2100 Agilent Bioanalyzer (Agilent Technologies, Inc). RNA-Seq libraries were prepared

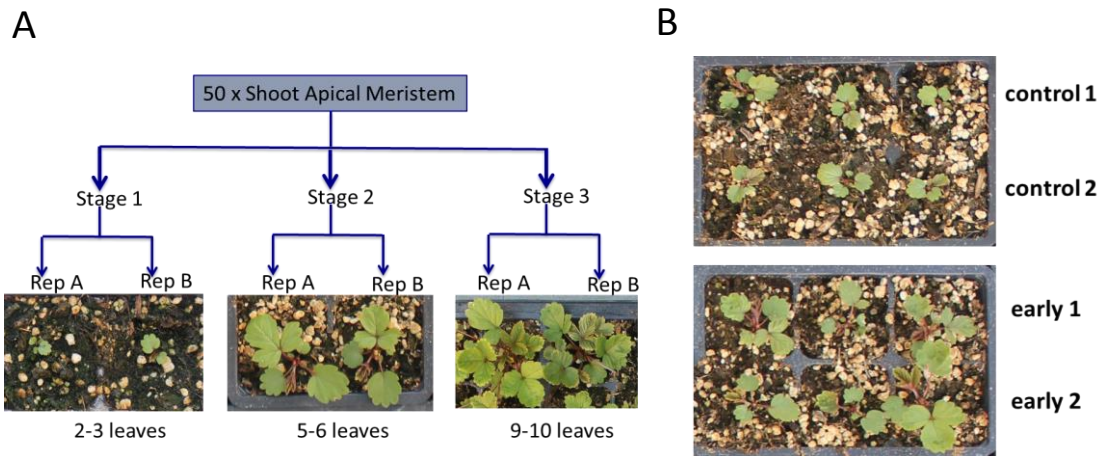


Figure 6.1 Shoot apical meristem sampling of early and control lines.
A: Experimental outline sampling shoot apical meristems from two independent control and early flowering lines, at three stages of development with two biological replicates from the S5 generation. Developmental stages 1, 2, 3 were sampled at 36, 50, and 64 days after sowing. B: Highlighting the difference in growth observed between control and early flowering lines 47 days after sowing. No difference in germination time was observed.

using TruSeq RNA sample prep v2 LS protocol (Illumina). All 24 libraries were sequenced on Illumina HiScanSQ platform according to the manufacturer's instructions.

6.3.3 Mapping and assembly of RNA-Seq reads

The raw reads were filtered and trimmed using Trimmomatic v0.32 (Bolger et al., 2014). The resulting FASTQ files were aligned to the reference woodland strawberry genome (https://www.rosaceae.org/species/fragaria/fragaria_vesca/genome_v1.0) using STAR 2.4.0 (Dobin et al., 2013).

6.3.4 Analysis of differential gene expression

FeatureCounts was used to quantify reads counts (Liao et al., 2014), and custom Perl script was used to combine all raw counts to form a data matrix. R package (version 3.2.1) of Bioconductor DESeq2 (Love et al., 2014) and edgeR version 3.10.5 were used to determine DEG (Robinson et al., 2010). Genes were retained if they were expressed in at least one biological replicate with greater than 1 read count per million (CPM) reads. After filtration of genes with low expression levels, the data were normalized using the default method trimmed mean of M values (TMM). Tagwise dispersions were estimated using the estimateDisp function. The Generalized Liner Model (GLM) likelihood ratio test was performed to test the differential expression for the following contrasts: stage 1 early versus stage 1 control; stage 2 early versus stage 2 control; stage 3 early versus stage 3 control. A multi-testing adjusted p-value (FDR) < 0.05 with log2-fold-change (logFC) >1 was set as cutoff for DEG.

6.3.5 DNA isolation, WGBS library preparation and sequencing

Genomic DNA was extracted from meristems using the Plant/Fungi DNA isolation Kit (Norgen, Thorold, ON, Canada). DNA quantification was performed using Qubit 2.0 Fluorometer and the Qubit DNA BR Assay Kits (Invitrogen, Eugene, OR, USA) according to the manufacturer's instructions. Genomic DNA (50ng) was treated with sodium bisulfite conversion reagent using the EZ DNA Methylation-Gold Kit (Catalog Nos. D5005, Zymo Research), followed by incubation in a thermal cycler at 98°C for 10 min, 64 °C for 2.5 h. The unmethylated Lambda DNA was used as a reference to calculate the conversion rate. After purification of modified DNA, the methylation libraries were prepared using a EpiGnome Methyl-Seq Kit protocol (Epicenter/Illumina). The quality of DNA methylation libraries were assessed via the high sensitivity DNA labchip on the 2100 Agilent Bioanalyzer

(Agilent Technologies, Inc). The libraries were sequenced on Illumina HiScanSQ platform according to the manufacturer's instructions.

6.3.6 Mapping and assembly of bisulfite sequencing reads

BSMAP was used to align reads to the reference strawberry genome (Xi and Li, 2009). Python script methratio.py was used to extract methylation ratios from the BSMAP aligned reads. To determine the confidence with which methylation state could be assigned at each cytosine a binomial test was performed for each cytosine in the *F. vesca* genome where a threshold of 0.05 was used as the cutoff. A custom perl script was used to generate cytosine methylation profiles in different genomic regions in the whole genome.

6.4 Results

6.4.1 Transcriptome profiles of SAM

In total, 24 libraries including four lines (control1, control2, early1, early2) with two biological replicates (A, B) in three stages (S1, S2, S3) were sequenced. The summary of Illumina pair end (PE) raw reads of these libraries characteristics are presented in **Table 6.1**. The raw reads ranged from 4,725,712 to 6,308,160 among these 24 libraries. A total number of 131,770,982 clean reads out of 134,199,428 raw reads (98.19%) were generated after filtering poor quality score reads and trimming adaptors. These clean reads were used in mapping to the reference genome and downstream analysis. We defined the requirement for expressed gene as these which were found in at least 12 libraries and the CPM was greater than one. Two replicates were used in these 24 libraries. In our study, we found expression for a total of 16,655 genes (48%) across the 24 libraries out of the 34,809 genes which were annotated in the *F. vesca* genome.

6.4.2 Differential gene expression study

The relationship among gene expression profiles was determined using hierarchical clustering (**Figure 6.2**). As anticipated, all eight stage 1 samples uniformly clustered together, suggesting the four control and four early samples were more related to each other at this stage. No obvious differential expression was observed at this stage. Differences in gene expression began to emerge at stage 2 where one of the control lines in stage 3 clustered with all stage 2 lines, indicating the similarity between stage 3 control and stage 2 control lines. Gene expression differences became more pronounced at stage 3. In particular, stage 3

Table 6.1 The summary of 24 libraries sequencing characteristics.

Sample	Raw reads	Clean reads	Surviving rate (%)
S1: control1A	5,063,949	4,959,782	97.94
S1: control1B	5,381,152	5,297,413	98.44
S1: control2A	5,340,665	5,224,136	97.82
S1: control2B	5,427,428	5,337,260	98.34
S1: early1A	5,460,778	5,340,623	97.8
S1: early1B	5,481,039	5,388,196	98.31
S1: early2A	4,796,771	4,702,161	98.03
S1: early2B	4,725,712	4,654,508	98.49
S2: control1A	6,016,873	5,898,816	98.04
S2: control1B	6,308,160	6,212,457	98.48
S2: control2A	5,883,294	5,767,127	98.03
S2: control2B	6,112,113	6,018,974	98.48
S2: early1A	5,914,532	5,796,186	98
S2: early1B	6,029,146	5,932,116	98.39
S2: early2A	5,324,488	5,213,964	97.92
S2: early2B	6,088,755	5,990,771	98.39
S3: control1A	5,086,877	4,977,963	97.86
S3: control1B	6,292,099	6,194,394	98.45
S3: control2A	5,969,784	5,848,013	97.96
S3: control2B	5,973,608	5,883,564	98.49
S3: early1A	5,465,755	5,354,187	97.96
S3: early1B	5,088,024	5,010,355	98.47
S3: early2A	5,493,788	5,382,854	97.98
S3: early2B	5,474,638	5,385,162	98.37
Total	134,199,428	131,770,982	98.19

Raw reads: reads generated by Illumina HiSeq 2000 platform.
Clean reads: reads after filtering poor quality score reads and trimming adaptors using Trimmomatic v0.32 program.
Surviving rate (%) = (Clean reads number / Raw reads number) * 100%
S1, S2, and S3 represent stage 1, stage 2, and stage 3. A and B are two biological replicates. For example S3:early2A means the stage 3 replicate A of early 2 lines. A total of 24 samples studied including three stages (S1, S2, S3), four lines (control 1, control 2, early 1, early 2) and two replicates (A, B).

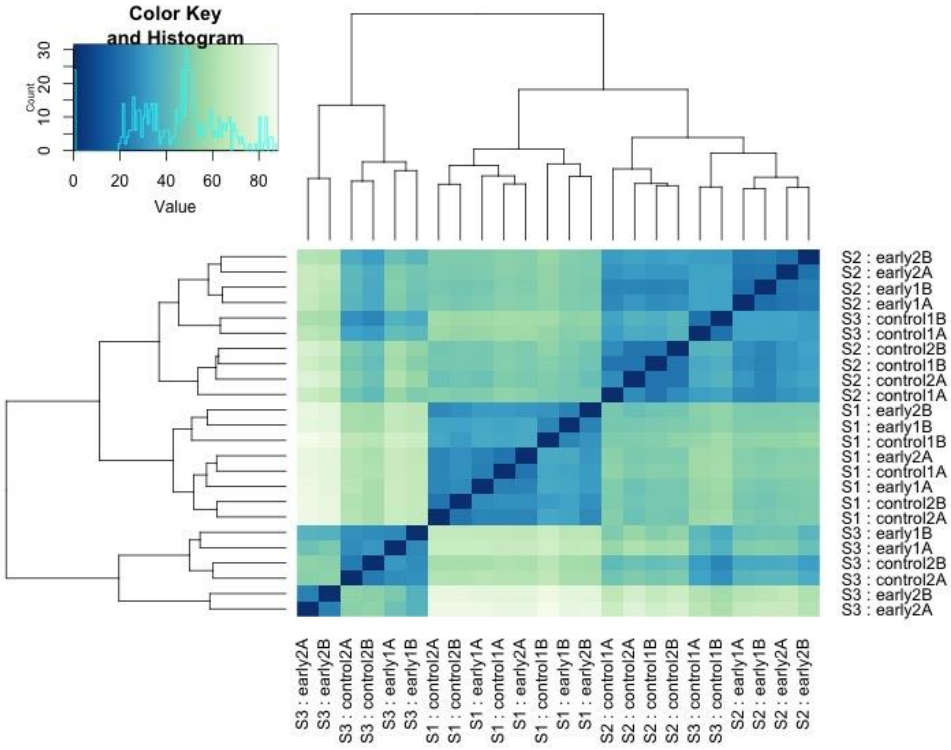


Figure 6.2 Clustering of transcriptome profiles among all examined 24 libraries.
S1, S2, and S3 represent stage 1, stage 2, and stage 3. A and B are two biological replicates. For example S3:early2A means the stage 3 replicate A of early 2 lines. A total of 24 samples studied including three stages (S1, S2, S3), four lines (control 1, control 2, early 1, early 2) and two replicates (A, B). The darker color with less value in color key and histogram indicated less sample to sample distance. All eight stage 1 samples clustered together as shown in the middle part of the heatmap. Differences in gene expression begun to emerge at stage 2 where stage 3 control lines clustered with stage 2 mutant lines as shown in the top right of heatmap. The remaining stage 3 lines clustered together as shown in the left bottom of heatmap.

early2A and stage 3 early2B (**Figure 6.2**) these two replicates clustered together and clearly separated from other stage 3 samples, suggesting early flowering mutants started to exhibit distinct differences in gene expression in stage 3.

To better understand the biological mechanism of early flowering, we conducted the DEG study between early and control lines among the three stages. A total of 1153 DEG were identified with 858 up-regulated and 295 down-regulated genes (**Figure 6.3**). The smallest number of DEG was detected in stage 1, with only five genes observed upregulated in early lines compared to control lines. In stage 2, there were a total of 132 and 85 DEG upregulated and downregulated, respectively. The majority of the differences occurred at stage 3. In stage 3, there were 931 DEG identified, with 721 upregulated and 210 downregulated. The functional annotations for the top five most DEG in stage 1 and the top ten most DEG in stage 2 and stage 3 were presented in **Table 6.2**. These DEG are involved in general metabolism, flavonoid metabolism and transcriptional activators with most of these genes are upregulated in early lines.

6.4.3 Expression of putative genes involved in the flowering time pathway

We further conducted a homologues search of the *Arabidopsis* flowering time genes in strawberry. The *F. vesca* genome contained 96 (55%) homologues of the 174 *Arabidopsis* genes involved in the transition to flowering and floral identity. The majority of these homologues are involved in vernalization and the autonomous pathway (40 genes) and photoperiod pathway (36 genes). Through the DNA sequencing of early 1 and early 2 lines in stage 3, and ERFv 153 line (the parental line from which early 1 and early 2 were derived) in Chapter 3, we verified no SNP existed among the 96 homologues genes controlling flowering time which indicates the early lines and control lines had identical genetic sequences in at least these flowering time genes. The further investigation of these 96 genes revealed six genes including *FUL*, (*CONSTANS LIKE 5*) *COL5*, *CIRCADIAN 1* (*CIR1*), *CDF1,3,5*, *LHY/CCA1*, *CO* with differential expression in stage 2, and six genes including *FT*, *PRR9*, *LATE MERISTEM IDENTITY 1* (*LMI1*), *SEP3*, *LHY*/*CCA1*, *NFYB2*/*HAP3B* with differential expression in stage3 (**Table 6.3**). Most of these genes were involved in the regulation of flowering in the photoperiod pathway. Flowering time inhibitors such as *CIR1*, *CDF1,3,5* were down-regulated in early flowering lines (**Table 6.3**). Since *CDFs* repress the expression of *CO*, this down-regulation contributes to early flowering (**Figure 6.4**). A number of genes

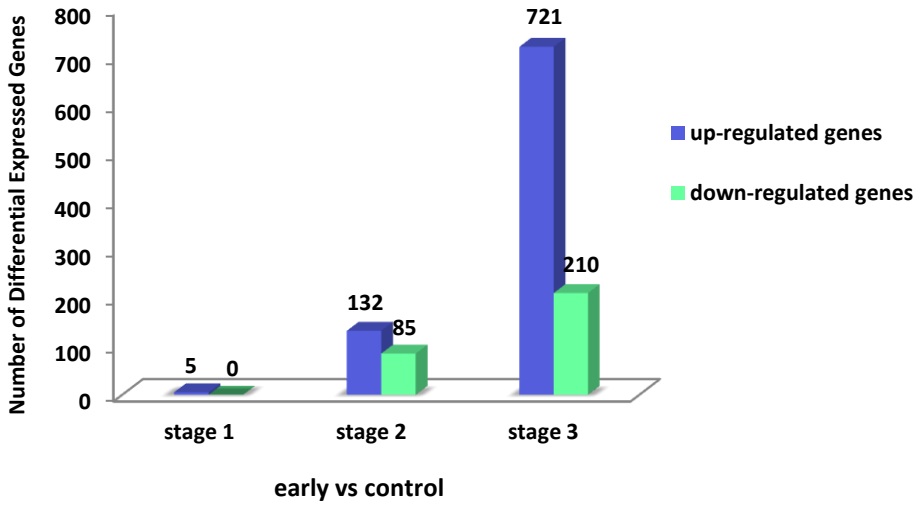


Figure 6.3 The number of differentially expressed genes detected between control and early flowering lines at the three stages (36, 50, and 64 days after sowing). The differentially expressed genes data of control 1 and control 2 was combined as control. The differential expressed genes data of early 1, early 2 was combined as early. The adjusted p-value False Discovery Rate (FDR) < 0.05 and log2-fold-change (logFC) > 1 were set as the cutoff for differentially expressed genes.

Table 6.2 The most differentially expressed genes in stage1, stage 2 and stage 3 in *F.vesca*.

Stage	Gene number	Gene ID	*Arabidopsis* homologues	Putative function	logFC	FDR
1	Fv2g19820	gene27467	no homologue	Putative_ribonuclease_H_protein_At1g65750_(probable)	5.60	2.31E-08
	Fv2g19840	gene27469	no homologue	(3R)-hydroxymyristoyl-[acyl-carrier-protein]_dehydratase_(probable)	1.97	1.35E-07
	Fv2g19810	gene27466	no homologue	Serine/threonine-protein_kinase_ATM_(AtATM)_(probable)	1.57	1.26E-07
	Fv2g19470	gene27433	no homologue	UDP-GalNAc:beta-1,_3-N-acetylgalactosaminyltransferase_2_(similar_to)	3.31	6.18E-07
	Fv5g01850	gene32393	AT5G48000	Cytochrome_P450_87A3_(putative)	1.21	1.18E-05
2	Fv1g16700	gene23949	no homologue	PR_domain_zinc_finger_protein_4_(probable)	1.40	7.43E-13
	Fv1g16690	gene23948	no homologue	Polygalacturonase_(PG),_Precursor_(probable)	1.88	4.16E-09
	Fv2g19810	gene27466	no homologue	Serine/threonine-protein_kinase_ATM_(AtATM)_(probable)	1.85	7.53E-09
	Fv2g37430	gene08631	AT1G42430	hypothetical_protein	-1.14	7.53E-09
	Fv5g21920	gene12192	AT1G49940	Kinesin_light_chain_4_(KLC_4)_(probable)	1.71	8.59E-09
	Fv5g16820	gene09396	no homologue	(RS)-norcoclaurine_6-O-methyltransferase_(6-OMT)_(probable)	-2.34	2.83E-08
	Fv3g19860	gene27141	no homologue	Structure-specific_endonuclease_subunit_SLX1_homolog_2_(similar_to)	-1.04	3.96E-08
	Fv7g34450	gene12417	AT3G49580	Stress_response_protein_NST1_(probable)	1.91	3.23E-07
	Fv2g19820	gene27467	no homologue	Putative_ribonuclease_H_protein_At1g65750_(probable)	3.97	4.59E-07
	Fv6g02350	gene31529	no homologue	Protein_UNUSUAL_FLORAL_ORGANS_(AtFBX1)_(probable)	4.58	4.64E-07
3	Fv7g35350	gene12773	no homologue	Probable_mitochondrial_chaperone_bcs1	2.26	6.17E-10
	Fv3g14660	gene29721	AT1G29340	U-box_domain-containing_protein_17_(putative)	1.35	6.17E-10
	Fv2g31380	gene02557	no homologue	LOB_domain-containing_protein_39_(AS2-like_protein_41)_(similar_to)	1.70	6.24E-10
	Fv4g19340	gene22995	no homologue	Mucin-2_(MUC-2),_Precursor_(probable)	1.55	6.24E-10
	Fv5g23090	gene12308	no homologue	Probable_WRKY_transcription_factor_11	1.71	1.32E-09
	Fv7g11920	gene03952	no homologue	TMV_resistance_protein_N_(probable)	2.68	1.35E-09
	Fv7g39090	gene13322	no homologue	Multidrug_and_toxin_extrusion_protein_2_(mMATE-2)_(probable)	1.90	1.35E-09
	Fv4g22500	gene00722	AT1G70170	Metalloendoproteinase_1,_Precursor_(similar_to)	3.31	1.45E-09
	Fv3g23470	gene03285	no homologue	Chitin-inducible_gibberellin-responsive_protein_1_(putative)	2.87	1.45E-09
	Fv1g29010	gene30503	AT5G37540	Aspartic_proteinase_nepenthesin-1,_Precursor_(probable)	2.15	1.45E-09

All five differentially expressed genes in stage 1 are listed. In stage 2 and stage 3, only the top ten most differentially expressed genes are listed. The adjusted p-value False Discovery Rate (FDR) < 0.05 and log2-fold-change (logFC) > 1 were set as the cutoff for differentially expressed genes.

Table 6.3 Differentially expressed genes in flowering time pathways in stage 2 and stage 3 in *F. vesca*.

Stage	Gene number	Gene ID	*Arabidopsis* homologues	Gene name in *Arabidopsis*	Flowering pathway	logFC	FDR
2	Fv5g11740	gene26119	AT5G60910	FUL (FRUITFUL)	Flower meristem identity	1.28	0.0015
	Fv4g10090	gene27383	AT5G57660	COL5 (CONSTANS LIKE 5)	Photoperiod	-1.18	0.0019
	Fv5g15780	gene09296	AT5G37260	CIR1 (CIRCADIAN 1)	photoperiod, circadian clock	-1.57	0.0122
	Fv5g13170	gene26022	AT1G69570	CDF5 (CYCLING DOF FACTOR 5)	Photoperiod	-1.53	0.0153
	Fv5g13170	gene26022	AT3G47500	CDF3 (CYCLING DOF FACTOR 3)	Photoperiod	-1.53	0.0153
	Fv5g13170	gene26022	AT5G62430	CDF1 (CYCLING DOF FACTOR 1)	Photoperiod	-1.53	0.0153
	Fv7g04140	gene18601	AT1G01060	LHY (LATE ELONGATED HYPOCOTYL)	photoperiod, circadian clock	-3.93	0.0170
	Fv7g04140	gene18601	AT2G46830	CCA1 (CIRCADIAN CLOCK ASSOCIATED 1)	photoperiod, circadian clock	-3.93	0.0170
	Fv6g41490	gene04172	AT5G15840	CO (CONSTANS)	Photoperiod	-1.79	0.0289
3	Fv6g00240	gene21535	AT1G65480	FT (FLOWERING LOCUS T)	Photoperiod	2.74	0.0000
	Fv7g04240	gene18611	AT2G46790	PRR9 (PSEUDO-RESPONSE REGULATOR 9)	photoperiod, circadian clock	-2.49	0.0016
	Fv6g19640	gene30097	AT5G03790	LMI1 (LATE MERISTEM IDENTITY 1)	Flower meristem identity	-1.16	0.0045
	Fv6g42070	gene04229	AT1G24260	SEP3 (SEPALLATA 3)	Flower development	1.07	0.0066
	Fv7g04140	gene18601	AT1G01060	LHY (LATE ELONGATED HYPOCOTYL)	photoperiod, circadian clock	-3.59	0.0115
	Fv7g04140	gene18601	AT2G46830	CCA1 (CIRCADIAN CLOCK ASSOCIATED 1)	photoperiod, circadian clock	-3.59	0.0115
	Fv2g42710	gene23541	AT5G47640	NFYB2/HAP3B	Photoperiod	1.21	0.0338

The adjusted p-value False Discovery Rate (FDR) < 0.05 and log2-fold-change (logFC) >1 were set as the cutoff for differentially expressed genes.

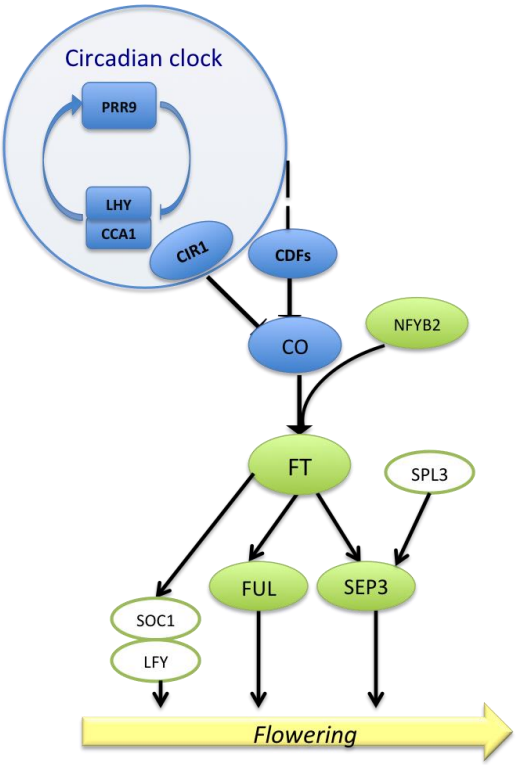


Figure 6.4 A simplified diagram showing the *F.vesca* homologues in the photoperiod pathway.

All of these genes listed are *F. vesca* homologues of *Arabidopsis.* The Up-regulated genes in early lines are in green; down-regulated genes in early lines are in blue. No gene expression changes are indicated by a green circle. For full gene names see list of abbreviations.

involved in the feedback loops regulating the circadian rhythm including *PRR9* and *LHY/CCA1* were also down-regulated in the early lines. Furthermore, the upregulation of downstream genes in the flowering pathway such as *FT*, *NFYB2/HAP3B*, *FUL* and *SEP3* promoted flowering transition and development to varying degrees. For example, the expression of flowering integrator *FT* was upregulated by more than six fold in the early flowering lines compared to the control suggesting a clear molecular basis for this phenotype. Other downstream genes (*NFYB2/HAP3B*, *FUL* and *SEP3*) were also upregulated in the early flowering lines by more than two fold (**Table 6.3**). *NFYB2/HAP3B* is a putative CCAAT-box binding transcription factor gene, and promotes flowering by activating the expression of *FT* and *SOC1* in *Arabidopsis* under a long day photoperiod (Cai et al., 2007). Floral meristem identity genes *FUL* and *SEP3* are the downstream target of *FT* (**Figure 6.4**).

6.4.4 DNA methylome profile in stage 3 meristem

Seventeen million cytosines were detected across the whole genome with a methylation rate of 17% in both control and early flowering lines determined using WGBS. The high resolution large scale WGBS enabled detection of methylated cytosines across all CG, CHG and CHH (where H = A, C, T) methylation sites. Sixty-two percent of the methylcytosines were in the CG context, 35% were in the CHG context, with the remaining 3% in the CHH context in both control and early flowering lines (**Figure 6.5 A, B**). Sixty-three percent of all CG, 36% of all CHG, and 1% of all CHH were methylated in control plants, which was similar to the corresponding level in early flowering lines, with 64% of all CG, 37% of all CHG and 1% of all CHH methylated (**Figure 6.5 C, D**). The MSAP methods of Chapter 3 can only recognize the 5'-CCGG-3' context based on the restriction enzymes digestion, and the methylated cytosines were 38% in control samples and thus the WGBS is a more comprehensive method to detect DNA methylation. Using WGBS in *Arabidopsis*, only 22% of all CG were methylated, which was only around one third of the early flowering plants (64%) in our study in *F. vesca* (Pradhan et al., 2008). Reports using WGBS in *Brassica oleracea* (Parkin et al., 2014) and rice (Feng et al., 2010) found 55% and 59% of all CG to be methylated respectively, and were comparable to our results.

6.4.5 DNA methylation and gene expression

The average CG methylation percentage in the promoter (upstream 1KB) and gene regions in control and early lines was explored. The gene expression data were normalized in edge R

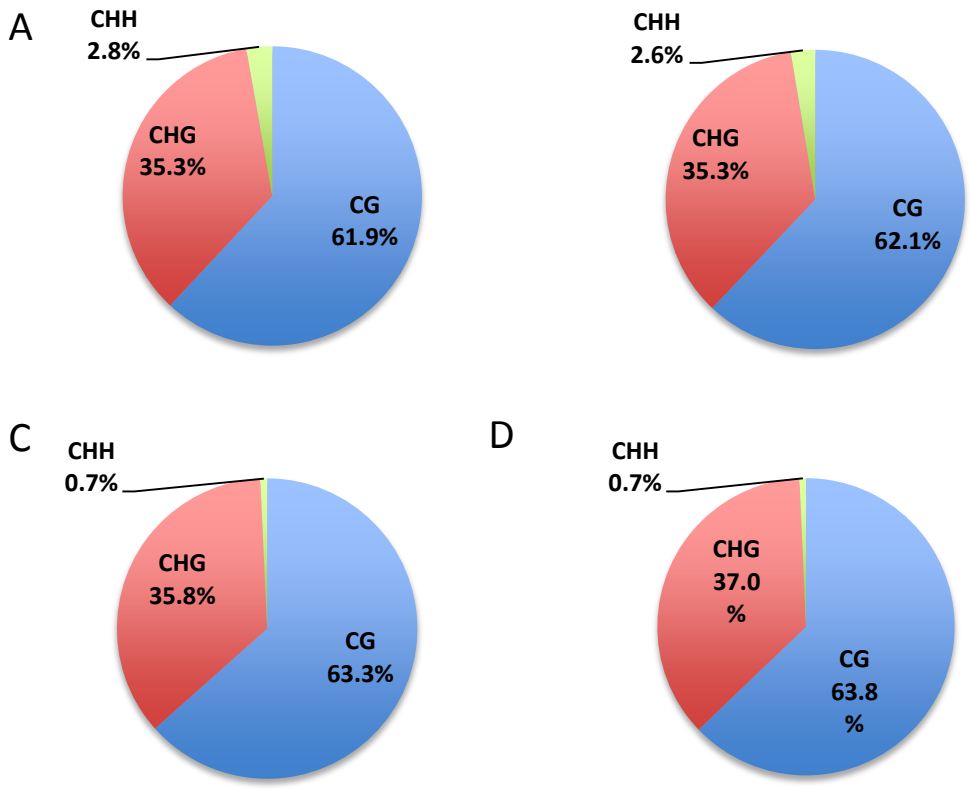


Figure 6.5 The distribution of methylcytosines among three sequence context CG, CHG, CHH.
A, B represent the percentage of methylated CG, CHG, and CHH of all methylated C in control and early flowering lines. C, D represent the percentage of methylated CG, CHG, and CHH of all CG, CHG, and CHH in control and early flowering lines.

using TMM, and log2 counts-per-million was used as its unit. There was no difference in the control and early lines, in distribution pattern when DNA methylation and gene expression were plotted in promoter and gene regions, therefore only the DNA methylation and gene expression in control lines were presented (**Figure 6.6**). CG methylation was not correlated with gene expression in the promoter region (**Figure 6.6 A**). This is in contrast to other studies which generally showed higher methylation in the promoter region was associated with lower gene expression. However, in the gene region itself, the more expressed genes tend to have a lower level of methylation (less than 25%) and the genes with moderate expression having variable levels of methylation (**Figure 6.6 B**).

6.4.6 DNA methylation of differentially expressed genes involved in the flowering time pathway

The methylation patterns of DEG including *FT*, *NFYB2*, *LHY*, *SEP3*, *LMI1*, *PRR9* in stage 3 were studied. Among these six gene regions tested, only *FT* and *LMI1* cytosine methylation in all of CG, CHG, and CHH DNA contexts were detected. CHH methylation was not found in the *SEP3* gene region and both CHG and CHH methylation were not observed in the *LHY* and *PRR9* gene regions. In the *NFYB2* gene region, the CHH methylation was only detected in the control lines but not in early flowering lines. The CG methylation level among replicates is less variable compared to CHG, and CHH methylation. Therefore, the detailed methylation level in gene regions of upstream 3KB, upstream 1KB, gene, coding sequence (CDS), intron, and downstream 3KB were evaluated in both control and early lines for CG methylation. Significant differences of CG methylation were detected in the *FT* gene and CDS regions between early and control lines ($P < 0.05$) (**Figure 6.7**). In control lines, the methylation level was localized by 41% within the gene, 9% higher than in the early lines. In contrast, early lines had 11% CG methylation in CDS while in control samples, the number was less than half (5%). This indicated most of the methylation in the gene exists in the intron region, with 62% in control lines and 51% in early lines. Differential methylation levels were observed across various gene regions in *NFYB2*, *LHY*, *SEP3*, *LMI1* and *PRR9*. The *NFYB2* gene region had the lowest methylation level, while the *LHY* gene region had the highest methylation level. However, apart from *FT* in the gene and CDS region, no significance was detected in control and early lines for methylation level in all gene regions in these five genes.

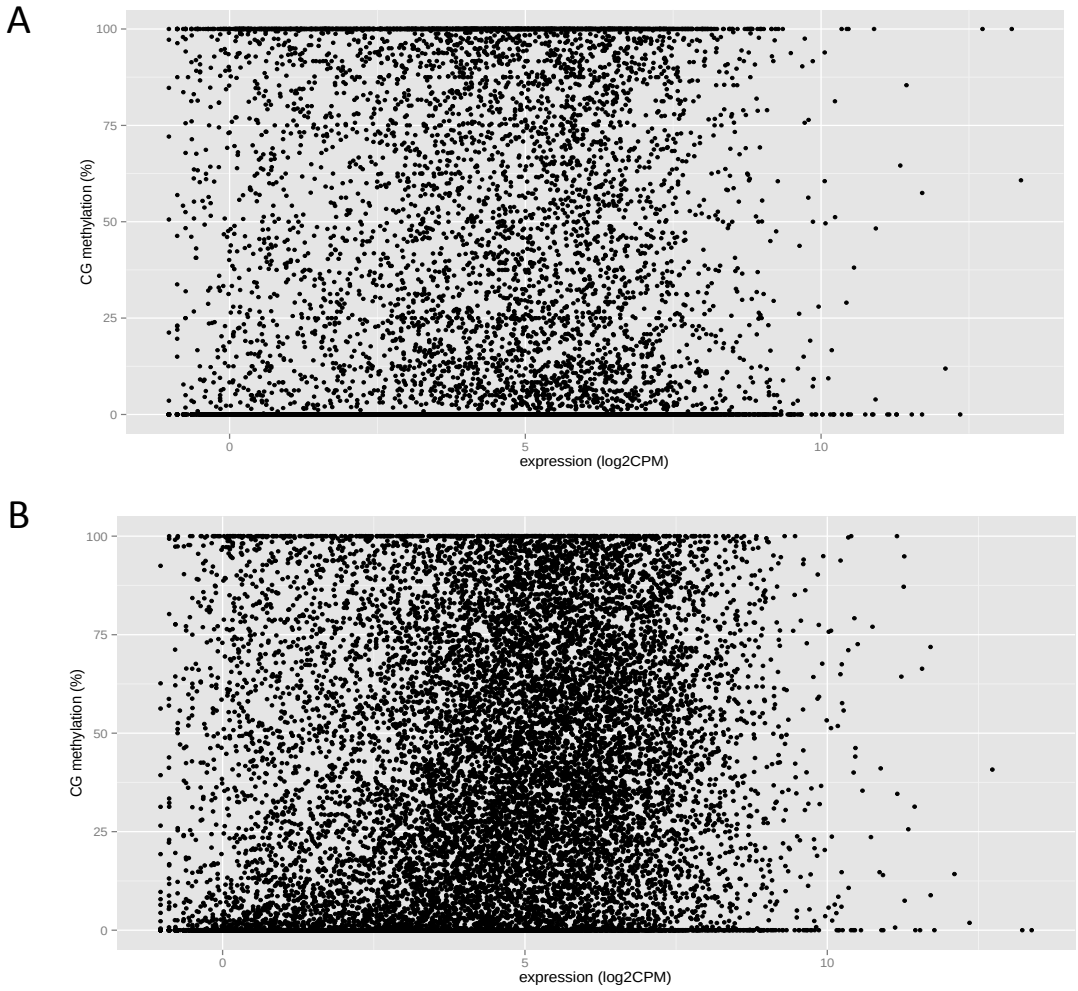


Figure 6.6 The relationship of DNA methylation and gene expression in promoter and gene regions.
A. Correlated of promoter (upstream 1KB) CG methylation and gene expression.
B. Correlated of gene CG methylation and gene expression.
X axis shows the percentage of CG methylation, y-axis shows gene expression value of log2 counts-per-million.

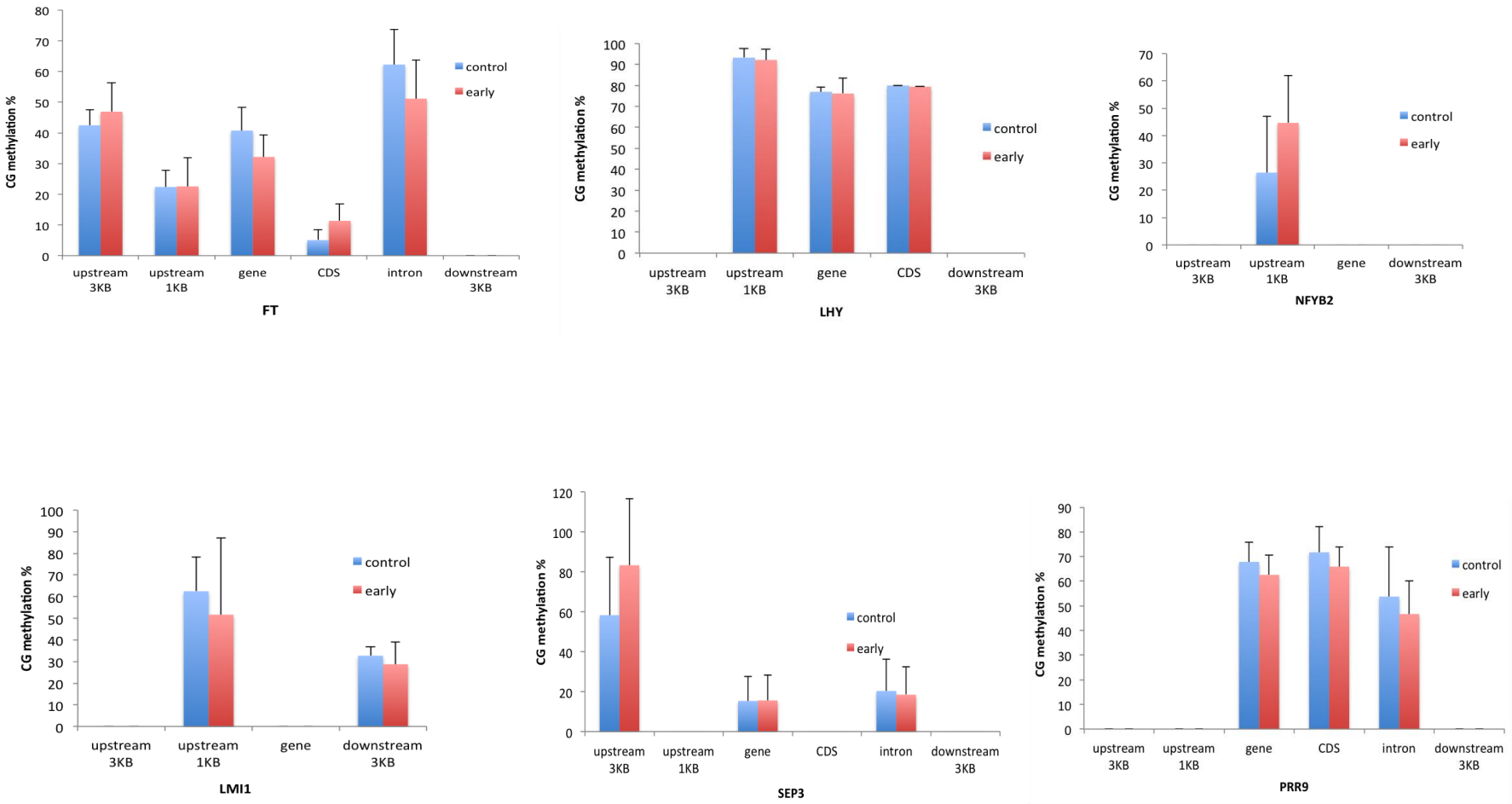


Figure 6.7 The level of CG methylation in differentially expressed flowering time genes in genomic regions. Significant differences were detected in the FT gene and CDS regions between early and control lines ($P < 0.05$).

6.5 Discussion

The early flowering material used in this experiment was selected from previous work, the fifth generation of an early flowering line ERFv153 derived from the 5-azaC treatment. The transgenerationally stable and heritable characteristics of this trait studied through previous chapters made it possible to further detect the molecular mechanisms underlying this trait. In this research we use three stages of meristem material to evaluate gene expression involved in floral transition in early flowering and control lines using RNA-Seq. The first stage was chosen at the 2-3 leaf stage based on a previous study which reported the expression of *AP1*, *LFY*, *SOC1* started to show large increases from the two leaf stage (Mouhu et al., 2009). The third stage was chosen around ten days before visible flowering at the 9-10 leaf stage in order to capture the expression of downstream genes in the floral transition.

A search for *F. vesca* homologues of the 174 *Arabidopsis* flowering time genes detected 96 genes in *F. vesca* covering all genetic pathways, with the majority of these genes being involved in the photoperiod, vernalization and the autonomous pathway. Other research identified 66 gene homologues of the 118 *Arabidopsis* flowering time genes using EST sequencing and was lacking few integrator genes (Mouhu et al., 2009). Most of these differentially expressed genes were in the photoperiod pathway. A total of six genes including *FUL*, *COL5*, *CIR1*, *CDF1*,*3*,*5*, *LHY/CCA1*, *CO* had differential expression in stage 2, and six genes including *FT*, *PRR9*, *LMI1*, *SEP3*, *LHY/CCA1*, *NFYB2/HAP3B* had differential expression in stage 3. *FT* is the integrator of the floral pathway and is regulated by photoperiod (Kardailsky et al., 1999). Studies in strawberry indicated *FT* is necessary for promotion of floral meristem identity genes expression (Ferrándiz et al., 2000; Koskela et al., 2012). In our study, there were six fold changes of *FT* attributed to the upregulation of floral meristem genes. The CCAAT-binding transcription factor *NFYB2/HAP3B* was verified to promote flowering through upregulating *FT* and *SOC1* in the photoperiod pathway in *Arabidopsis* (Cai et al., 2007). The study of *NFYB2/HAP3B* homologue in *F. vesca* has not been previously reported to the best of our knowledge. Apart from the function in carpel and fruit development, the MADS box transcription factor *FUL* is also one of the floral meristem identity genes activated by *FT*, promoting flowering at later stages (Gu et al., 1998; Calonje et al., 2004; Koskela et al., 2012; Torti et al., 2012). Similarly, another MADS box transcription factor *SEP3* is also regulated by *FT* and accumulates in meristems contributing to the floral transition (Teper-Bamnolker and Samach, 2005). The upregulation of downstream genes *FUL*

and *SEP3* in our early lines is directly associated with earlier flowering time compared to control lines.

In this chapter we detected gene expression profiles using RNA-seq in meristems in both early and control lines. WGBS as a method to comprehensively study DNA methylation at CG, CHG, and CHH contexts was used in stage 3 DNA methylation detection. It allows mapping the DNA methylation to the genome and more accurately detecting methylation profiles in detail. We found a similar level in the overall methylation rate at CG, CHG, CHH sites in early and control lines, with 63% of all CG, 36% of all CHG, and 1% of all CHH sites were methylated. Global genomic DNA methylation was quantified in strawberry (*F.* ×*ananassa*) using HPLC and found the methylation level ranged from 20-30% in young leaves (Zhang et al., 2012). Our research is the first report to comprehensively detect the methylation levels in *F. vesca* in the meristem. The overall DNA methylation in CG, CHG, and CHH in early lines and control lines using WGBS does not show differences. However, this does not mean a lack of DNA methylation variations between early lines and control lines since the loss and gain of methylation across the whole genome can render the net methylation level unchanged (Zhang et al., 2009).

Numerous studies in the past have explored the correlation between gene expression changes and DNA methylation levels and patterns (Zilberman et al., 2007; Li et al., 2012; Lokk et al., 2014; Parkin et al., 2014). Previous studies showed DNA methylation of the gene body is mainly limited to CG sites rather than all CG, CHG, and CHH sites (Zhang et al., 2006). Therefore, in our study, the impact of CG methylation on gene expression in genes and promoters was explored. The results of the moderate level of expressed genes show variable levels of methylation in gene regions and is similar to the study in *Arabidopsis* which found genes with moderate transcription level tend to be methylated (Zilberman et al., 2007). However, no direct relationship was observed between CG methylation and gene expression, which is different from other work in which high level of methylation in promoter areas was associated with low gene expression.

In *Arabidopsis*, the promoter region contains two direct repeats leading to methylation of the *FLOWERING WAGENIGEN* (*FWA*) gene, decreased methylation was associated with the FWA transcriptional activation (Soppe et al., 2000). Not only does the methylation in the promoter region affect gene expression, but also the gene expression of transcribed regions is

upregulated as a result of demethylation even if it is not in the promoter region (Zilberman et al., 2007). Here we further detect the methylation profiles of flowering time pathway DEG in different regions including upstream, downstream, gene region, and intron of the genome. There were large differences of the methylation profiles in CG, CHG, and CHH among the six genes *FT*, *NFYB2*, *LHY*, *SEP3*, *LMI1*, *PRR9* in stage 3. Stably heritable short clusters of CG methylation within the gene body were detected (Tran et al., 2005). The variation of CHG and CHH methylation between replicates might indicate the unstable inheritance of these sites and they tend to be more susceptible to change. We detected the variable CG methylation profiles in different regions of these six genes, suggesting the unique methylation profiles of these DEG in meristems. However, only significant differences were observed in the *FT* gene and *FT* CDS regions between early and control lines. The lack of widespread significant differences between early lines and control lines might indicate that CG methylation changes are not a direct reason contributing to genes expression variations and there might be other factors involved in this process.

6.6 Conclusions

In this study, the stable early flowering lines after five generations of selection were used to detect the molecular bases underlying the early flowering phenotype. The RNA-Seq analysis of meristem indicated the differences of DEG in three stages, with significant gene expression changes in stage 2 and stage 3. The homologue search of *Arabidopsis* flowering time genes (174) indicated 96 genes in *F. vesca* flowering pathway with the majority involved in vernalization and the autonomous pathway (40 genes), and photoperiod pathway (36 genes). There were six DEG in stage 2 and six DEG in stage 3 in *F. vesca* flowering pathway. The *FT* in stage 3 showed the highest fold change (six fold), which was likely the major contributor to the early flowering phenotype. The investigation of the stage 3 methylome indicated the total methylation levels of CG, CHG, CHH in *F. vesca* SAM. Further detection of methylation of DEG identified the significantly different CG methylation in the gene (early 32%, control 41%) and coding sequence regions (early 11%, control 5%) of *FT* in early flowering lines compared to control lines. This study provided insight into the molecular basis including changes in gene expression and DNA methylation in early flowering lines compared to control lines. Future research will focus on crossing between early lines and control lines to map the epialleles related to the early flowering trait.

7.0 GENERAL DISCUSSION AND CONCLUSIONS

Epigenetics has been a hot topic in the past decades with much research conducted on mammals, plants, and fungi. Plants provide ideal model systems to study epigenetics: 1) unlike mammals where DNA methylation is mainly limited to the CG site, in plants DNA methylation also occurs at CHG and CHH sites, guided by small RNA in *de novo* methylation at CHH site; 2) plants can tolerate null mutations in many pathways and chromatin regulators, while in mammals the loss is lethal (Committee, 2012); 3) in plants, transgenerational inheritance of acquired epigenetic status/epigenetic traits are common, while this is rare in mammals (Morgan et al., 1999; Rakyan et al., 2003). A number of epigenetic phenomena in many plants have been reported, either naturally occurring (Cubas et al., 1999; Chandler and Stam, 2004; Manning et al., 2006) or artificially induced (Sano et al., 1990; Fieldes, 1994b). The study areas include epigenetic control of plant growth and development, paramutations and transposable elements, stress adaptation, ecology and evolution. Most of the above research was conducted in model plant *Arabidopsis* or field crops with short life cycle of sexual reproduction. In the Rosaceae family, most members are heterozygous trees and shrubs with complex genetic background, and long seed to seed life cycles which are disadvantageous to investigate epigenetic mechanisms and inheritance studies. Thus, the comprehensive studies of epigenetic regulation and inheritance in the Rosaceae family have not been widely reported, although some epigenetic responses have been examined in apple (Li et al., 2002; Kumar et al., 2016), cherry (Avramidou et al., 2015) and rose (Ma et al., 2015).

Fragaria vesca is an ancestral diploid strawberry and a model species for the *Fragaria* genus as well as the Rosaceae family (Folta and Davis, 2006; Shulaev et al., 2008). *F. vesca* has the advantages of short generation time (3.5 month seed to seed), propagation through both seeds (meiosis) and runners (mitosis). Compared to other species in the *Fragaria* genus, *F. vesca* has a relatively small genome size (240MB). Although there are many strawberry species with a series of ploidy levels and wide distribution, the cultivated strawberry genetics base is still perceived to be very narrow (Debnath et al., 2008; Höfer et al., 2012). Therefore, using methods beyond traditional genetics to create new resources with broader phenotypic variation within *F. vesca* will help to enrich the germplasm for breeding. We used *F. vesca* to generate new resources of an epimutagenized population by treating seeds with 5-azaC, a DNA demethylating reagent.

The generated new data addressed fundamental questions of epigenetic regulation of phenotypic traits, the inheritance of selected traits through meiosis and mitosis in the model system *F. vesca* in the Rosaceae family. For the first time, the DNA methylation levels in wild type *F. vesca* were investigated and characterized. In addition, although most literature recognize 5-azaC as a plant epimutagen without base sequence changes (Holliday and Ho, 2002), induction of mutagenic changes has been reported (Katoh et al., 1993). However, there is a lack of clear investigation of this potential mutagenic effect of 5-azaC. Our research used two methods Amplified Fragment Length Polymorphism (AFLP) and DNA sequencing to survey the genetic background of the epimutagenized population and control population. The AFLP marker has very high reproducibility and can be used in a genetic diversity study (Jones et al., 1997; Mueller and Wolfenbarger, 1999). This method has been used widely in surveying the genetic background of a population before DNA methylation is studied (Marfil et al., 2009; Zhang et al., 2009; Hanai et al., 2010; Gao et al., 2016). In this research, the genetic identical background of *F. vesca* control and treated seedlings based on DNA sequencing was observed. This verified there is no introgression of alleles or genetic mutation induced by 5-azaC treatment, which made it possible to continue the epigenetic study with minimal influence from the genetic variation.

In previous studies, DNA methylation in strawberry was limited to qualitative analysis (Hao et al., 2002; Martelli et al., 2008) or quantitative calculation of DNA methylation changes in dormancy induction using High Performance Liquid Chromatography (HPLC) in strawberry cultivars (Zhang et al., 2012). However, the HPLC method to study DNA methylation only results in assessing the global DNA methylation level without knowledge of methylation patterns on cytosine. By contrast, the DNA methylation level in *F. vesca* was assayed using both Methylation Sensitive Amplified Polymorphisms (MSAP) and whole genome bisulfite sequencing (WGBS) in our research. The MSAP has the advantage of surveying the methylation level in a population with lower cost and more output information about the methylation patterns of recognition sites. The WGBS is a gold standard method to study DNA methylation at single nucleotide resolution. This is the first report in *F. vesca* using these two methods to study DNA methylation. The MSAP method exploited isoschizomers *Hp*a II and *Msp* I which recognize 5'-CCGG-3' and identified 38% methylation level in wide type. The WGBS is based on single nucleotide resolution in which a total of 63% of all CG, 36% of all CHG, and 1% of all CHH were methylated.

The most distinct phenotypic traits which responded to 5-azaC treatment in our study were flowering time, rosette diameter and stolon emergence time. Similar to flowering time, plant height and plant growth in *Arabidopsis* (Johannes et al., 2009; Reinders et al., 2009), the continuous distribution of flowering time, rosette diameter and stolon emergence time were observed in our research. These morphological characteristics are all important traits in breeding and are polygenically controlled, and affected by genetic and environmental conditions (Heide, 1977). The lack of genetic diversity is the main limitation in conventional breeding and crop adaptation to new environments. Epigenetic variation inducing phenotypic changes can be a potential strategy to modify phenotypic traits for crop improvement through careful selection of novel epialleles related to the desired traits as well as adaptation to the environment (Springer, 2013). Compared to the control population, an expanded variation of these three phenotypic traits was observed under 5-azaC. The extreme variants of both directions of the flowering time trait were observed. By contrast, in rosette diameter and stolon emergence time, only one tail of the small rosette diameter and late stolon emergence time extreme variants were observed. This might be a result of the differing numbers of genes controlling these quantitative traits, with 5-azaC randomly targeting on a small subset of loci across the genome resulting in genome wide stochastic epialleles. These epialleles can give rise to different visual characters of plants if they are related to the regulation of the phenotypic traits (Zhang and Hsieh, 2013). The following inheritance and selection study demonstrated that the epialleles that control the phenotypic variance, especially the early flowering trait, can be heritable through meiosis and subject to selection.

The molecular bases underlying the heritable early flowering lines were detected through both standard and novel MSAP experiment (Chapter 5). We found the symmetric CG and CHG methylation sites were maintained in the early flowering lines, rather than the CHH asymmetric methylation site. The mechanisms of DNA methylation maintenance at CG, CHG, and CHH were different (Cao and Jacobsen, 2002; Kankel et al., 2003; Henderson and Jacobsen, 2007). Methylation at symmetric CG and CHG sites were maintained during each cycle of DNA replication in cell division based on template maintenance by DNA METHYLTRANSFERASE 1 (MET1) and CHROMOMETHYLASE (CMT3). By contrast, CHH methylation was maintained through *de novo* methylation catalyzed by DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), with small RNA targeted at genomic DNA sequences, termed RNA-directed DNA methylation (RdDM). In addition, the

de novo DNA methylation can also establish methylation at CG and CHG sequence context. This might suggest the early flowering trait was related to methylation pattern changes at CG and CHG sites. When the DNA methylation pathway are inhibited by 5-azaC, the DNA methylation at CG, CHG sites are diluted after DNA replication, called demethylation (He et al., 2011). The demethylation status at double-stranded symmetric CG, CHG can be maintained in the following cycles of DNA replication by MET1 and CMT3. Therefore, the demethylation status at CG and CHG was able to be accumulated in early flowering lines to form a distinct methylation status as indicated by the clearly separated epigenetic profiles in **Figure 3.10** in Chapter 3 using the standard MSAP method which recognizes 5'-CCGG-3' (CG, CHG). The epialleles causing the late flowering trait might also be related to demethylation at CG and CHG sites since we detected significantly different epigenetic profiles in late flowering lines compared to control lines as indicated in **Figure 3.10** in Chapter 3. Unlike the total separation of the early flowering trait, there was overlap between control lines and late flowering lines which might suggest some of the demethylation status at CG and CHG revert back to methylation status by *de novo* DNA methylation. This may explain why the late flowering trait was not able to transmit to the offspring, indicating the epigenetic trait was reversible, and not as stable as mutations which modify the DNA sequence.

In plant development, flowering is an important stage with four pathways participating in this process. Both endogenous and environmental factors are involved in the flowering pathways, with environmental factors comprising regulatory factors of the photoperiod and temperature (Boss et al., 2004). The environmental factors were likely to be affected by natural selection while the remaining cues were conservative. Early flowering tends to be naturally selected in the environment with short or non-favorable growing conditions (Roux et al., 2006). The material used in our study is *F. vesca*, which has a wider distribution than other *Fragaria* species, throughout north temperate Europe, America, Asia, North Africa and South America and a long growing history. By the 1300s, *F. vesca* was transplanted from the wilderness to gardens by the French (Darrow, 1966). From an evolutionary and adaptation perspective, *F. vesca* would have had to adjust its flowering time in order to colonize more widely. Photoperiod is one of the major environmental cues regulating flowering time. Our study indicated the stable inheritance in early flowering lines induced by a DNA demethylating reagent and the subsequent RNA-Seq study found most of the differentially expressed flowering time genes were in the photoperiod pathway. Thus, it appears that the epialleles

related to the photoperiod pathway are likely to be the targets of selection. Since epigenetic variation plays an important role in ecology and evolutionary processes (Bossdorf et al., 2008; Richards et al., 2010), our finding might provide a new insight into the evolutionary response to selection of the early flowering trait in *F. vesca*. From the perspective of evolution, compared to the unstable late flowering, maybe the epialleles control early flowering were more conserved through evolution than late flowering trait in order to finish the life cycle in adaptation.

This research used the model plant *F. vesca* ssp. *vesca* accession Hawaii 4 (germplasm accession: PI551572) in the Rosaceae family to explore the epigenetic responses initiated by inhibiting DNA methylation using the demethylating reagent, 5-azaC. The hypomethylated epimutagenized population possessed a broader range of several quantitative traits and established an important resource for inheritance and selection studies. Finally the fifth generation early flowering lines were characterized by RNA-Seq and WGBS to provide the molecular background for this inherited phenotypic trait. This research provided a fundamental understanding of the potential application of epigenetic regulated phenotypic traits for crop improvement. The conclusions were:

1. A genetically uniform population resource with 305 5-azaC treatment lines and 59 control lines was developed. The expanded phenotypic variation of several quantitative traits such as flowering time, rosette diameter, and stolon emergence time were identified.

2. Early flowering and late stolon emergence traits can be successfully transmitted into subsequent generations and subject to selection. The stable transmission of the early flowering trait through both meiosis and mitosis was demonstrated.

3. Novel MSAP method was developed to detect methylation at CG, CHG, and CHH, complementing the standard MSAP. The symmetric CG, CHG methylation was found to be maintained over generations in the early flowering lines, whereas the asymmetric CHH methylation pattern was not maintained over generations.

4. A total of 96 homologues of the *Arabidopsis* flowering time genes were found in *F. vesca*. Most of the differentially expressed genes involved in floral transition were in

the photoperiod pathway. The up-regulation of downstream genes in flowering pathway *NFYB2/HAP3B*, *FT*, *FUL* and *SEP3* which promote flowering was observed in early flowering lines. Significantly different CG methylation levels in the gene and coding sequence regions of *FT* in early and control lines was observed.

5. Cytosine methylation level was detected in *F. vesca*, MSAP method based on 5'-CCGG-3' indicated 38% cytosine methylation level, whereas WGBS detected a total of 63% of all CG, 36% of all CHG, and 1% of all CHH were methylated.

In future, a population segregating for flowering time will be produced by generating selfed-seeds from a cross of the early flowering with control line to map the early flowering epialleles. WGBS of extremes in the flowering time distribution through Bulked Segregant Analysis (BSA) can be used to locate the underlying epialleles.

8.0 REFERENCES

Akimoto, K., Katakami, H., Kim, H.J., Ogawa, E., Sano, C.M., Wada, Y., and Sano, H. (2007). Epigenetic inheritance in rice plants. Ann Bot **100,** 205-217.

Albani, M., Battey, N., and Wilkinson, M. (2004). The development of ISSR-derived SCAR markers around the SEASONAL FLOWERING LOCUS (SFL) in Fragaria vesca. Theoretical and applied genetics **109,** 571-579.

Amasino, R. (2010). Seasonal and developmental timing of flowering. The Plant Journal **61,** 1001-1013.

Amasino, R.M., and Michaels, S.D. (2010). The timing of flowering. Plant Physiol **154,** 516-520.

Amoah, S., Kurup, S., Rodriguez Lopez, C.M., Welham, S.J., Powers, S.J., Hopkins, C.J., Wilkinson, M.J., and King, G.J. (2012). A hypomethylated population of Brassica rapa for forward and reverse epi-genetics. BMC Plant Biol **12,** 193.

Andrés, F., and Coupland, G. (2012). The genetic basis of flowering responses to seasonal cues. Nature Reviews Genetics **13,** 627-639.

Angers, B., Castonguay, E., and Massicotte, R. (2010). Environmentally induced phenotypes and DNA methylation: how to deal with unpredictable conditions until the next generation and after. Mol Ecol **19,** 1283-1295.

Archbold, D., and Dennis Jr, F. (1985). Strawberry receptacle growth and endogenous IAA content as affected by growth regulator application and achene removal. Journal of the American Society for Horticultural Science (USA).

Arnau, G., Lallemand, J., and Bourgoin, M. (2000). Are AFLP markers the best alternative for cultivar identification? In International Symposium on Molecular Markers for Characterizing Genotypes and Identifying Cultivars in Horticulture 546, pp. 301-306.

Avramidou, E.V., Ganopoulos, I.V., Doulis, A.G., Tsaftaris, A.S., and Aravanopoulos, F.A. (2015). Beyond population genetics: natural epigenetic variation in wild cherry (Prunus avium). Tree Genetics & Genomes **11,** 1-9.

Bastow, R., Mylne, J.S., Lister, C., Lippman, Z., Martienssen, R.A., and Dean, C. (2004). Vernalization requires epigenetic silencing of FLC by histone methylation. Nature **427,** 164-167.

Bateson, W., and Pellew, C. (1915). On the genetics of "rogues" among culinary peas (Pisum sativum). Journal of Genetics **5,** 13-36.

Bauer, A. (1992). Progress in breeding decaploid Fragaria× vescana. In II International Strawberry Symposium 348, pp. 60-64.

Beck, S., and Rakyan, V.K. (2008). The methylome: approaches for global DNA methylation profiling. Trends in genetics : TIG **24,** 231-237.

Becker, C., and Weigel, D. (2012). Epigenetic variation: Origin and transgenerational inheritance. Current Opinion in Plant Biology **15,** 562-567.

Becker, C., Hagmann, J., Müller, J., Koenig, D., Stegle, O., Borgwardt, K., and Weigel, D. (2011). Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature **480,** 245-249.

Bednarek, P.T., Orlowska, R., Koebner, R.M., and Zimny, J. (2007). Quantification of the tissue-culture induced variation in barley (Hordeum vulgare L.). BMC Plant Biol **7,** 10.

Bell, J.T., Pai, A.A., Pickrell, J.K., Gaffney, D.J., Pique-Regi, R., Degner, J.F., Gilad, Y., and Pritchard, J.K. (2011). DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol **12,** R10.

Bell, J.T., Tsai, P.-C., Yang, T.-P., Pidsley, R., Nisbet, J., Glass, D., Mangino, M., Zhai, G., Zhang, F., and Valdes, A. (2012). Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. PLoS Genet **8,** e1002629.

Bender, J. (2004). DNA methylation and epigenetics. Annu Rev Plant Biol **55,** 41-68.

Berg, A., Meza, T.J., Mahić, M., Thorstensen, T., Kristiansen, K., and Aalen, R.B. (2003). Ten members of the Arabidopsis gene family encoding methyl‐CpG‐binding domain proteins are transcriptionally active and at least one, AtMBD11, is crucial for normal development. Nucleic acids research **31,** 5291-5304.

Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. Journal of Molecular Biology **203,** 971-983.

Bewley, J.D. (1997). Seed germination and dormancy. The plant cell **9,** 1055.

Bilichak, A., Ilnystkyy, Y., Hollunder, J., and Kovalchuk, I. (2012). The progeny of Arabidopsis thaliana plants exposed to salt exhibit changes in DNA methylation, histone modifications and gene expression. PloS one **7,** e30515.

Bird, A. (1992). The essentials of DNA methylation. Cell **70,** 5-8.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes and Development **16,** 6-21.

Bird, A.P., and Southern, E.M. (1978). Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from Xenopus laevis. Journal of molecular biology **118,** 27-47.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics**,** btu170.

Bors, R.H. (2000). A STREAMLINED SYNTHETlC OCTOPLOID SYSTEM THAT EMPHASIZES FRAGARIA VESCA AS A BRIDGE SPEClES (The University of Guelph).

Boss, P.K., Bastow, R.M., Mylne, J.S., and Dean, C. (2004). Multiple pathways in the decision to flower: enabling, promoting, and resetting. The Plant Cell **16,** S18-S31.

Bossdorf, O. (2010). Experimental alteration of DNA methylation affects the phenotypic plasticity of ecologically relevant traits in Arabidopsis thaliana. Evolutionary Ecology **24,** 541.

Bossdorf, O., Richards, C.L., and Pigliucci, M. (2008). Epigenetics for ecologists. Ecology letters **11,** 106-115.

Bossdorf, O., Arcuri, D., Richards, C.L., and Pigliucci, M. (2010). Experimental alteration of DNA methylation affects the phenotypic plasticity of ecologically relevant traits in Arabidopsis thaliana. Evolutionary Ecology **24,** 541-553.

Bringhurst, R.S. (1990). Cytogenetics and evolution in American Fragaria. HortScience **25,** 879-881.

Broday, L., Lee, Y.-W., and Costa, M. (1999). 5-azacytidine induces transgene silencing by DNA methylation in Chinese hamster cells. Molecular and cellular biology **19,** 3198-3204.

Brown, J.C.L., De Decker, M.M., and Fieldes, M.A. (2008). A comparative analysis of developmental profiles for DNA methylation in 5-azacytidine-induced early-flowering flax lines and their control. Plant Science **175,** 217-225.

Brown, T., and Wareing, P. (1965). The genetical control of the everbearing habit and three other characters in varieties of Fragaria vesca. Euphytica **14,** 97-112.

Burkhardt, R.W. (1979). The Spirit of System: Lamarck and Evolutionary Biology.

Burn, J.E., Bagnall, D.J., Metzger, J.D., Dennis, E.S., and Peacock, W.J. (1993). DNA methylation, vernalization, and the initiation of flowering. Proceedings of the National Academy of Sciences of the United States of America **90,** 287-291.

Cai, X., Ballif, J., Endo, S., Davis, E., Liang, M., Chen, D., DeWald, D., Kreps, J., Zhu, T., and Wu, Y. (2007). A putative CCAAT-binding transcription factor is a regulator of flowering timing in Arabidopsis. Plant Physiology **145,** 98-105.

Callahan, H., and Pigliucci, M. (2005). Indirect consequences of artificial selection on plasticity to light quality in Arabidopsis thaliana. Journal of evolutionary biology **18,** 1403-1415.

Calonje, M., Cubas, P., Martínez-Zapater, J.M., and Carmona, M.J. (2004). Floral meristem identity genes are expressed during tendril development in grapevine. Plant Physiology **135,** 1491-1501.

Cao, X., and Jacobsen, S.E. (2002). Role of the Arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. Current Biology **12,** 1138-1144.

Cedar, H., Solage, A., Glaser, G., and Razin, A. (1979). Direct detection of methylated cytosine in DNA by use of the restriction enzyme Mspl. Nucleic acids research **6,** 2125-2132.

Chakrabarty, D., Yu, K.-W., and Paek, K.Y. (2003). Detection of DNA methylation changes during somatic embryogenesis of Siberian ginseng (Eleuterococcus senticosus). Plant Science **165,** 61-68.

Chandler, V.L., and Stam, M. (2004). Chromatin conversations: mechanisms and implications of paramutation. Nature Reviews Genetics **5,** 532-544.

Chatterjee, M., Bermudez-Lozano, C.L., Clancy, M.A., Davis, T.M., and Folta, K.M. (2011). A strawberry KNOX gene regulates leaf, flower and meristem architecture. PLoS One **6,** e24752.

Chen, M., Lv, S., and Meng, Y. (2010). Epigenetic performers in plants. Development Growth and Differentiation **52,** 555-566.

Childers, N. (1981). The Strawberry Cultivars to Marketing. (Horticultural Publ.).

Chouard, P. (1960). Vernalization and its relations to dormancy. Annual Review of Plant Physiology **11,** 191-238.

Christman, J.K. (2002). 5 - Azacytidine and 5 - aza - 2' - deoxycytidine as inhibitors of DNA methylation : mechanistic studies and their implications for cancer therapy. Oncogene **21,** 5483.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly **6,** 80-92.

Clark, S.J., Statham, A., Stirzaker, C., Molloy, P.L., and Frommer, M. (2006). DNA methylation: bisulphite modification and analysis. Nature protocols **1,** 2353-2364.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature **452,** 215-219.

Committee, E.P. (2012). Reading the second code: Mapping epigenomes to understand plant growth, development, and adaptation to the environment. The Plant Cell **24,** 2257-2261.

Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S., Gissot, L., and Turnbull, C. (2007). FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. science **316,** 1030-1033.

Crossa, J., Campos Gde, L., Perez, P., Gianola, D., Burgueno, J., Araus, J.L., Makumbi, D., Singh, R.P., Dreisigacker, S., Yan, J., Arief, V., Banziger, M., and Braun, H.J. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics **186,** 713-724.

Cubas, P., Vincent, C., and Coen, E. (1999). An epigenetic mutation responsible for natural variation in floral symmetry. Nature **401,** 157-161.

Dalakouras, A., Dadami, E., Zwiebel, M., Krczal, G., and Wassenegger, M. (2012). Transgenerational maintenance of transgene body CG but not CHG and CHH methylation. Epigenetics **7,** 1071-1078.

Darnell, R.L., Cantliffe, D.J., Kirschbaum, D.S., and Chandler, C.K. (2003). The physiology of flowering in strawberry. Hortic. Rev **28,** 325-349.

Darrow, G.M. (1936). Interrelation of temperature and photoperiodism in the production of fruit-buds and runners in the strawberry. In Proc. Amer. Soc. Hort. Sci, pp. 363.

Darrow, G.M. (1966). The strawberry. History, breeding and physiology. The strawberry. History, breeding and physiology.

Darwish, O., Shahan, R., Liu, Z., Slovin, J.P., and Alkharouf, N.W. (2015). Re-annotation of the woodland strawberry (Fragaria vesca) genome. BMC genomics **16,** 1.

Davis, T., and Yu, H. (1997). A linkage map of the diploid strawberry, Fragaria vesca. Journal of Heredity **88,** 215-221.

Debnath, S., Khanizadeh, S., Jamieson, A., and Kempler, C. (2008). Inter Simple Sequence Repeat (ISSR) markers to assess genetic diversity and relatedness within strawberry genotypes. Canadian journal of plant science **88,** 313.

Degani, C., Rowland, L.J., Saunders, J.A., Hokanson, S.C., Ogden, E.L., Golan-Goldhirsh, A., and Galletta, G.J. (2001). A comparison of genetic relationship measures in strawberry (Fragaria× ananassa Duch.) based on AFLPs, RAPDs, and pedigree data. Euphytica **117,** 1-12.

Diez, C.M., Roessler, K., and Gaut, B.S. (2014). Epigenetics and plant genome evolution. Current opinion in plant biology **18,** 1-8.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics **29,** 15-21.

Dreher, T., and Poovaiah, B. (1982). Changes in auxin content during development in strawberry fruits. In HORTSCIENCE (AMER SOC HORTICULTURAL SCIENCE 701 NORTH SAINT ASAPH STREET, ALEXANDRIA, VA 22314-1998), pp. 475-475.

Dunlap, J.C. (1999). Molecular bases for circadian clocks. Cell **96,** 271-290.

Durner, E.F. (1984). Photoperiod and temperature effects on flower and runner development in day-neutral, Junebearing, and everbearing strawberries. J. Amer. Soc. Hort. Sci. **109,** 396-400.

Dwivedi, R.S., Herman, J.G., McCaffrey, T.A., and Raj, D.S. (2011). Beyond genetics: epigenetic code in chronic kidney disease. Kidney international **79,** 23-32.

Eichten, S.R., Briskine, R., Song, J., Li, Q., Swanson-Wagner, R., Hermanson, P.J., Waters, A.J., Starr, E., West, P.T., Tiffin, P., Myers, C.L., Vaughn, M.W., and

Springer, N.M. (2013). Epigenetic and genetic influences on DNA methylation variation in maize populations. Plant Cell **25,** 2783-2797.

Fait, A., Hanhineva, K., Beleggia, R., Dai, N., Rogachev, I., Nikiforova, V.J., Fernie, A.R., and Aharoni, A. (2008). Reconfiguration of the achene and receptacle metabolic networks during strawberry fruit development. Plant physiology **148,** 730-750.

Falke, K.C., Glander, S., He, F., Hu, J., de Meaux, J., and Schmitz, G. (2013). The spectrum of mutations controlling complex traits and the genetics of fitness in plants. Current opinion in genetics & development **23,** 665-671.

Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., Yanofsky, M.F., Kater, M.M., and Colombo, L. (2003). MADS-box protein complexes control carpel and ovule development in Arabidopsis. The Plant Cell **15,** 2603-2611.

Feil, R., and Fraga, M.F. (2011). Epigenetics and the environment: emerging patterns and implications. Nat Rev Genet **13,** 97-109.

Feng, S., and Jacobsen, S.E. (2011). Epigenetic modifications in plants: an evolutionary perspective. Current opinion in plant biology **14,** 179-186.

Feng, S., Jacobsen, S.E., and Reik, W. (2010). Epigenetic reprogramming in plant and animal development. Science **330,** 622-627.

Ferrándiz, C., Gu, Q., Martienssen, R., and Yanofsky, M.F. (2000). Redundant regulation of meristem identity and plant architecture by FRUITFULL, APETALA1 and CAULIFLOWER. Development **127,** 725-734.

Fieldes, M.A. (1994a). An examination of the β - glucosidase (linamarase) banding pattern in flax seedlings using Ferguson plots and sodium dodecyl sulphate - polyacrylamide gel electrophoresis. Electrophoresis **15,** 654.

Fieldes, M.A. (1994b). Heritable effects of 5-azacytidine treatments on the growth and development of flax (Linum usitatissimum) genotrophs and genotypes. Genome **37,** 1-11.

Fieldes, M.A., and Amyot, L.M. (1999). Epigenetic control of early flowering in flax lines induced by 5- azacytidine applied to germinating seed. Journal of Heredity **90,** 199-206.

Fieldes, M.A., Schaeffer, S.M., Krech, M.J., and Brown, J.C. (2005). DNA hypomethylation in 5-azacytidine-induced early-flowering lines of flax. Theor Appl Genet **111,** 136-149.

Finnegan, E.J. (1996). The role of DNA methylation in plant development. Cold Spring Harbor Monograph Archive **32,** 127-140.

Finnegan, E.J., Peacock, W.J., and Dennis, E.S. (1996). Reduced DNA methylation in Arabidopsis thaliana results in abnormal plant development. Proceedings of the National Academy of Sciences **93,** 8449-8454.

Finnegan, E.J., Sheldon, C.C., Jardinaud, F., Peacock, W.J., and Dennis, E.S. (2004). A cluster of Arabidopsis genes with a coordinate response to an environmental stimulus. Current Biology **14,** 911-916.

Folta, K.M., and Davis, T.M. (2006). Strawberry genes and genomics. Critical Reviews in Plant Sciences **25,** 399-415.

Fornara, F., de Montaigu, A., and Coupland, G. (2010). SnapShot: control of flowering in Arabidopsis. Cell **141,** 550-550. e552.

Forster, B.P., and Thomas, W.T. (2005). Doubled haploids in genetics and plant breeding. Plant Breed Rev **25,** 57-88.

Friml, J. (2003). Auxin transport—shaping the plant. Current opinion in plant biology **6,** 7-12.

Fujimoto, R., Sasaki, T., Ishikawa, R., Osabe, K., Kawanabe, T., and Dennis, E.S. (2012). Molecular mechanisms of epigenetic variation in plants. Int J Mol Sci **13,** 9900-9922.

Fulneček, J., and Kovařík, A. (2014). How to interpret Methylation Sensitive Amplified Polymorphism (MSAP) profiles? BMC Genetics **15**.

Galvão, A.G., Resende, L.V., Guimaraes, R.M., Ferraz, A.K.L., Morales, R.G.F., Marodin, J.C., and Catão, H.C.R.M. (2014). Overcoming strawberry achene dormancy for improved seedling production in breeding programs. Idesia **32,** 57-62.

Gao, L., Geng, Y., Li, B., Chen, J., and Yang, J. (2010). Genome-wide DNA methylation alterations of Alternanthera philoxeroides in natural and manipulated habitats: implications for epigenetic regulation of rapid responses to environmental fluctuation and phenotypic variation. Plant Cell Environ **33,** 1820-1827.

Gao, L., Diarso, M., Zhang, A., Zhang, H., Dong, Y., Liu, L., Lv, Z., and Liu, B. (2016). Heritable alteration of DNA methylation induced by whole‐chromosome aneuploidy in wheat. New Phytologist **209,** 364-375.

Gidoni, D., Rom, M., Kunik, T., Zur, M., Izsak, E., Izhar, S., and Firon, N. (1994). Strawberry - cultivar Identification using Randomly Amplified Polymorphic DNA (RAPD) Markers. Plant Breeding **113,** 339-342.

Gillaspy, G., Ben-David, H., and Gruissem, W. (1993). Fruits: a developmental perspective. The Plant Cell **5,** 1439.

Gitan, R.S., Shi, H., Chen, C.M., Yan, P.S., and Huang, T.H.M. (2002). Methylation-specific oligonucleotide microarray: A new potential for high-throughput methylation analysis. Genome research **12,** 158-164.

Gocal, G.F., Sheldon, C.C., Gubler, F., Moritz, T., Bagnall, D.J., MacMillan, C.P., Li, S.F., Parish, R.W., Dennis, E.S., and Weigel, D. (2001). GAMYB-like genes, flowering, and gibberellin signaling in Arabidopsis. Plant Physiology **127,** 1682-1693.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annu Rev Biochem **74,** 481-514.

González, R.M., Ricardi, M.M., and Iusem, N.D. (2011). Atypical epigenetic mark in an atypical location: Cytosine methylation at asymmetric (CNN) sites within the body of a non-repetitive tomato gene. BMC Plant Biology **11**.

Graham, J., McNicol, R., and McNicol, J. (1996). A comparison of methods for the estimation of genetic diversity in strawberry cultivars. Theoretical and Applied Genetics **93,** 402-406.

Grant-Downton, R.T., and Dickinson, H. (2006). Epigenetics and its implications for plant biology 2. The 'epigenetic epiphany': epigenetics, evolution and beyond. Annals of Botany **97,** 11-27.

Greenup, A., Peacock, W.J., Dennis, E.S., and Trevaskis, B. (2009). The molecular biology of seasonal flowering-responses in Arabidopsis and the cereals. Annals of Botany**,** mcp063.

Grewal, S.I., and Jia, S. (2007). Heterochromatin revisited. Nature Reviews Genetics **8,** 35-46.

Gu, Q., Ferrándiz, C., Yanofsky, M.F., and Martienssen, R. (1998). The FRUITFULL MADS-box gene mediates cell differentiation during Arabidopsis fruit development. Development **125,** 1509-1517.

Hanai, L.R., Floh, E.I.S., Fungaro, M.H.P., Santa-Catarina, C., de Paula, F.M., Viana, A.M., and Vieira, M.L.C. (2010). Methylation patterns revealed by MSAP profiling

in genetically stable somatic embryogenic cultures of Ocotea catharinensis (Lauraceae). In Vitro Cellular & Developmental Biology-Plant **46,** 368-377.

Hao, Y.-J., You, C.-X., and Deng, X.-X. (2002). Analysis of ploidy and the patterns of amplified fragment length polymorphism and methylation sensitive amplified polymorphism in strawberry plants recovered from cryopreservation. CryoLetters **23,** 37-46.

Hartmann, H. (1947). The influence of temperature on the photoperiodic response of several strawberry varieties grown under controlled environment conditions. In Proceedings of the American Society for Horticultural Science (AMER SOC HORTICULTURAL SCIENCE 701 NORTH SAINT ASAPH STREET, ALEXANDRIA, VA 22314-1998), pp. 243-245.

Hashimshony, T., Zhang, J., Keshet, I., Bustin, M., and Cedar, H. (2003). The role of DNA methylation in setting up chromatin structure during development. Nature Genet. **34,** 187-192.

Hauben, M., Haesendonckx, B., Standaert, E., Van Der Kelen, K., Azmi, A., Akpo, H., Van Breusegem, F., Guisez, Y., Bots, M., and Lambert, B. (2009). Energy use efficiency is characterized by an epigenetic component that can be directed through artificial selection to increase yield. Proceedings of the National Academy of Sciences **106,** 20109-20114.

He, X.-J., Chen, T., and Zhu, J.-K. (2011). Regulation and function of DNA methylation in plants and animals. Cell research **21,** 442-465.

He, Y. (2012). Chromatin regulation of flowering. Trends in plant science **17,** 556-562.

Heard, E., and Martienssen, R.A. (2014). Transgenerational epigenetic inheritance: Myths and mechanisms. Cell **157,** 95-109.

Heide, O.M. (1977). Photoperiod and temperature interactions in growth and flowering of strawberry. Physiologia Plantarum **40,** 21-26.

Hempel, F.D., and Feldman, L.J. (1994). Bi-directional inflorescence development inArabidopsis thaliana: Acropetal initiation of flowers and basipetal initiation of paraclades. Planta **192,** 276-286.

Henderson, I.R., and Jacobsen, S.E. (2007). Epigenetic inheritance in plants. Nature **447,** 418-424.

Henderson, I.R., Chan, S.R., Cao, X., Johnson, L., and Jacobsen, S.E. (2010). Accurate sodium bisulfite sequencing in plants. Epigenetics **5,** 47-49.

Herrera, C.M., and Bazaga, P. (2010). Epigenetic differentiation and relationship to adaptive genetic divergence in discrete populations of the violet Viola cazorlensis. New Phytol **187,** 867-876.

Hetzl, J., Foerster, A.M., Raidl, G., and Scheid, O.M. (2007). CyMATE: a new tool for methylation analysis of plant genomic DNA after bisulphite sequencing. The Plant Journal **51,** 526-536.

Hirsch, S., Baumberger, R., and Grossniklaus, U. (2012). Epigenetic variation, inheritance, and selection in plant populations. In Cold Spring Harbor symposia on quantitative biology (Cold Spring Harbor Laboratory Press), pp. 97-104.

Höfer, M., Drewes-Alwarez, R., Scheewe, P., and Olbricht, K. (2012). Morphological evaluation of 108 strawberry cultivars–and consequences for the use of descriptors. Journal of Berry Research **2,** 191-206.

Hollender, C.A., Geretz, A.C., Slovin, J.P., and Liu, Z. (2012). Flower and early fruit development in a diploid strawberry, Fragaria vesca. Planta **235,** 1123-1139.

Hollender, C.A., Kang, C., Darwish, O., Geretz, A., Matthews, B.F., Slovin, J., Alkharouf, N., and Liu, Z. (2014). Floral transcriptomes in woodland strawberry uncover developing receptacle and anther gene networks. Plant physiology **165,** 1062-1075.

Holliday, R. (1994). Epigenetics: an overview. Developmental genetics **15,** 453-457.

Holliday, R., and Ho, T. (2002). DNA methylation and epigenetic inheritance. Methods **27,** 179-183.

Hu, Y., Morota, G., Rosa, G.J., and Gianola, D. (2015). Prediction of Plant Height in Arabidopsis thaliana Using DNA Methylation Data. Genetics **201,** 779-793.

Huang, T., Böhlenius, H., Eriksson, S., Parcy, F., and Nilsson, O. (2005). The mRNA of the Arabidopsis gene FT moves from leaf to shoot apex and induces flowering. Science **309,** 1694-1696.

Hummer, K.E., and Janick, J. (2009). Rosaceae: taxonomy, economic importance, genomics. In Genetics and genomics of Rosaceae (Springer), pp. 1-17.

Hwan Lee, J., Joon Kim, J., and Ahn, J.H. (2012). Role of SEPALLATA3 (SEP3) as a downstream gene of miR156-SPL3-FT circuitry in ambient temperature-responsive flowering. Plant signaling & behavior **7,** 1151-1154.

Hytönen, T., Palonen, P., Mouhu, K., and Junttila, O. (2004). Crown branching and cropping potential in strawberry (Fragaria ananassa Duch.) can be enhanced by

daylength treatments. The Journal of Horticultural Science and Biotechnology **79,** 466-471.

Ietswaart, R., Wu, Z., and Dean, C. (2012). Flowering time control: another window to the connection between antisense RNA and chromatin. Trends in Genetics **28,** 445-453.

Jablonka, E., and Raz, G. (2009). Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. The Quarterly review of biology **84,** 131-176.

Jaeger, K.E., and Wigge, P.A. (2007). FT protein acts as a long-range signal in Arabidopsis. Current Biology **17,** 1050-1054.

Jahn, O.L., and Dana, M.N. (1970). Crown and inflorescence development in the strawberry, Fragaria ananassa. American Journal of Botany**,** 605-612.

Johannes, F., Colot, V., and Jansen, R.C. (2008). Epigenome dynamics: a quantitative genetics perspective. Nature Reviews Genetics **9,** 883-890.

Johannes, F., Porcher, E., Teixeira, F.K., Saliba-Colombani, V., Simon, M., Agier, N., Bulski, A., Albuisson, J., Heredia, F., Audigier, P., Bouchez, D., Dillmann, C., Guerche, P., Hospital, F., and Colot, V. (2009). Assessing the impact of transgenerational epigenetic variation on complex traits. PLoS Genetics **5**.

Johansson, M., and Staiger, D. (2015). Time to flower: interplay between photoperiod and the circadian clock. Journal of experimental botany **66,** 719-730.

Johnson, H.W., Robinson, H., and Comstock, R. (1955). Estimates of genetic and environmental variability in soybeans. Agronomy journal **47,** 314-318.

Johnston, J.W., Harding, K., Bremner, D.H., Souch, G., Green, J., Lynch, P.T., Grout, B., and Benson, E.E. (2005). HPLC analysis of plant DNA methylation: a study of critical methodological factors. Plant Physiol Biochem **43,** 844-853.

Jones, C., Edwards, K., Castaglione, S., Winfield, M., Sala, F., Van de Wiel, C., Bredemeijer, G., Vosman, B., Matthes, M., and Daly, A. (1997). Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Molecular breeding **3,** 381-390.

Jones, P.A., Taylor, S.M., and Wilson, V.L. (1983). Inhibition of DNA methylation by 5-azacytidine. Recent results in cancer research. Fortschritte der Krebsforschung. Progres dans les recherches sur le cancer **84,** 202-211.

Jordan, C.Y., Ally, D., and Hodgins, K.A. (2015). When can stress facilitate divergence by altering time to flowering? Ecology and evolution **5,** 5962-5973.

Jung, S., Ficklin, S.P., Lee, T., Cheng, C.-H., Blenda, A., Zheng, P., Yu, J., Bombarely, A., Cho, I., and Ru, S. (2014). The genome database for rosaceae (GDR): year 10 update. Nucleic acids research **42,** D1237-D1244.

Jüttermann, R., Li, E., and Jaenisch, R. (1994). Toxicity of 5-aza-2'-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. Proceedings of the National Academy of Sciences **91,** 11797-11801.

Kakutani, T. (1997). Genetic characterization of late‐flowering traits induced by DNA hypomethylation mutation in Arabidopsis thaliana. The plant journal **12,** 1447-1451.

Kakutani, T., Munakata, K., Richards, E.J., and Hirochika, H. (1999). Meiotically and mitotically stable inheritance of DNA hypomethylation induced by ddm1 mutation of Arabidopsis thaliana. Genetics **151,** 831-838.

Kankel, M.W., Ramsey, D.E., Stokes, T.L., Flowers, S.K., Haag, J.R., Jeddeloh, J.A., Riddle, N.C., Verbsky, M.L., and Richards, E.J. (2003). Arabidopsis MET1 cytosine methyltransferase mutants. Genetics **163,** 1109-1122.

Kardailsky, I., Shukla, V.K., Ahn, J.H., Dagenais, N., Christensen, S.K., Nguyen, J.T., Chory, J., Harrison, M.J., and Weigel, D. (1999). Activation tagging of the floral inducer FT. Science **286,** 1962-1965.

Karlić, R., Chung, H.-R., Lasserre, J., Vlahoviček, K., and Vingron, M. (2010). Histone modification levels are predictive for gene expression. Proceedings of the National Academy of Sciences **107,** 2926-2931.

Katoh, Y., Maekawa, M., and Sano, Y. (1993). Effects of 5-azacytidine on somatic mutation in a soybean test system. Mutation Research/Genetic Toxicology **300,** 49-55.

Kawashima, T., and Berger, F. (2014). Epigenetic reprogramming in plant sexual reproduction. Nature Reviews Genetics **15,** 613-624.

Kim, D.-H., Doyle, M.R., Sung, S., and Amasino, R.M. (2009). Vernalization: winter and the timing of flowering in plants. Annual Review of Cell and Developmental **25,** 277-299.

Kim, J.-M., To, T.K., Ishida, J., Morosawa, T., Kawashima, M., Matsui, A., Toyoda, T., Kimura, H., Shinozaki, K., and Seki, M. (2008). Alterations of lysine modifications on the histone H3 N-tail under drought stress conditions in Arabidopsis thaliana. Plant and Cell Physiology **49,** 1580-1588.

Koornneef, M., Alonso-Blanco, C., Peeters, A.J., and Soppe, W. (1998). Genetic control of flowering time in Arabidopsis. Annual review of plant biology **49,** 345-370.

Koornneef, M., Vries, H., Hanhart, C., Soppe, W., and Peeters, T. (1994). The phenotype of some late‐flowering mutants is enhanced by a locus on chromosome 5 that is not effective in the Landsberg erecta wild‐type. The Plant Journal **6,** 911-919.

Koskela, E.A., Mouhu, K., Albani, M.C., Kurokura, T., Rantanen, M., Sargent, D.J., Battey, N.H., Coupland, G., Elomaa, P., and Hytönen, T. (2012). Mutation in TERMINAL FLOWER1 reverses the photoperiodic requirement for flowering in the wild strawberry Fragaria vesca. Plant Physiology **159,** 1043-1054.

Kou, H.P., Li, Y., Song, X.X., Ou, X.F., Xing, S.C., Ma, J., Von Wettstein, D., and Liu, B. (2011). Heritable alteration in DNA methylation induced by nitrogen-deficiency stress accompanies enhanced tolerance by progenies to the stress in rice (Oryza sativa L.). Journal of Plant Physiology **168,** 1685-1693.

Krueger, F., Kreck, B., Franke, A., and Andrews, S.R. (2012). DNA methylome analysis using short bisulfite sequencing data. Nature Methods **9,** 145-151.

Kumar, G., Rattan, U.K., and Singh, A.K. (2016). Chilling-Mediated DNA Methylation Changes during Dormancy and Its Release Reveal the Importance of Epigenetic Regulation during Winter Dormancy in Apple (Malus x domestica Borkh.). PloS one **11,** e0149934.

Kuo, K.C., McCune, R.A., Gehrke, C., Midgett, R., and Ehrlich, M. (1980). Quantitative reversed-phase high performance liquid chromatographic determination of major and modified deoxyribonucleosides in DNA. Nucleic acids research **8,** 4763-4776.

Kurokura, T., Mimida, N., Battey, N.H., and Hytönen, T. (2013). The regulation of seasonal flowering in the Rosaceae. Journal of experimental botany **64,** 4131-4141.

Lamkey, K., and Lee, M. (1993). Quantitative genetics, molecular markers, and plant improvement. Focused Plant Improvement**,** 18-23.

Lauria, M., Piccinini, S., Pirona, R., Lund, G., Viotti, A., and Motto, M. (2014). Epigenetic variation, inheritance, and parent-of-origin effects of cytosine methylation in maize (Zea mays). Genetics **196,** 653-666.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet **11,** 204-220.

Lee, J., and Lee, I. (2010). Regulation and function of SOC1, a flowering pathway integrator. Journal of Experimental Botany **61,** 2247-2254.

Levy, Y.Y., and Dean, C. (1998). The transition to flowering. Plant Cell **10,** 1973-1989.

Li, X., Xu, M., and Korban, S.S. (2002). DNA methylation profiles differ between field-andin vitro-grown leaves of apple. Journal of Plant Physiology **159,** 1229-1234.

Li, X., Zhu, J., Hu, F., Ge, S., Ye, M., Xiang, H., Zhang, G., Zheng, X., Zhang, H., and Zhang, S. (2012). Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. BMC genomics **13,** 300.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics **30,** 923-930.

Lindroth, A.M., Cao, X., Jackson, J.P., Zilberman, D., McCallum, C.M., Henikoff, S., and Jacobsen, S.E. (2001). Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science **292,** 2077-2080.

Lippman, Z., May, B., Yordan, C., Singer, T., and Martienssen, R. (2003). Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. PLoS Biol **1,** e67.

Lokk, K., Modhukur, V., Rajashekar, B., Märtens, K., Mägi, R., Kolde, R., Koltšina, M., Nilsson, T.K., Vilo, J., and Salumets, A. (2014). DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. Genome biology **15,** 1.

López-Aranda, J.M., Soria, C., Santos, B.M., Miranda, L., Domínguez, P., and Medina-Mínguez, J.J. (2011). Strawberry production in mild climates of the world: a review of current cultivar use. International journal of fruit science **11,** 232-244.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **15,** 550.

Ma, N., Chen, W., Fan, T., Tian, Y., Zhang, S., Zeng, D., and Li, Y. (2015). Low temperature-induced DNA hypermethylation attenuates expression of RhAG, an AGAMOUS homolog, and increases petal number in rose (Rosa hybrida). BMC plant biology **15,** 1.

Mandel, M.A., and Yanofsky, M.F. (1995). A gene triggering flower formation in Arabidopsis. Nature **377,** 522-524.

Manning, K., Tor, M., Poole, M., Hong, Y., Thompson, A.J., King, G.J., Giovannoni, J.J., and Seymour, G.B. (2006). A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nat Genet **38,** 948-952.

Mardis, E.R. (2008). The impact of next-generation sequencing technology on genetics. Trends in genetics **24,** 133-141.

Marfil, C.F., Camadro, E.L., and Masuelli, R.W. (2009). Phenotypic instability and epigenetic variability in a diploid potato of hybrid origin, Solanum ruiz-lealii. BMC Plant Biol **9,** 21.

Marfil, C.F., Asurmendi, S., and Masuelli, R.W. (2012). Changes in micro RNA expression in a wild tuber-bearing Solanum species induced by 5-Azacytidine treatment. Plant Cell Rep **31,** 1449-1461.

Martelli, G., Zaccagnino, N., Milella, L., and Greco, I. (2008). Characterization of the DNA methylation activity by gene expression analysis in Fragaria vesca. In VI International Strawberry Symposium 842, pp. 569-572.

Martienssen, R.A., and Colot, V. (2001). DNA methylation and epigenetic inheritance in plants and filamentous fungi. Science **293,** 1070-1074.

Martin, C., and Zhang, Y. (2007). Mechanisms of epigenetic inheritance. Current opinion in cell biology **19,** 266-272.

Mayer, A.M., and Poljakoff-Mayber, A. (1982). The Germination of Seeds: Pergamon International Library of Science, Technology, Engineering and Social Studies. (Elsevier).

McClelland, M., Nelson, M., and Raschke, E. (1994). Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. Nucleic Acids Research **22,** 3640-3659.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., and Daly, M. (2010). The Genome

Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research **20,** 1297-1303.

Melo, W.M.C., Balestre, M., Von Pinho, R.G., and de Sousa Bueno Filho, J.S. (2014). Genetic control of the performance of maize hybrids using complex pedigrees and microsatellite markers. Euphytica **195,** 331-344.

Meyer, P., Niedenhof, I., and ten Lohuis, M. (1994). Evidence for cytosine methylation of non-symmetrical sequences in transgenic Petunia hybrida. The EMBO journal **13,** 2084.

Michael, and Walsh, L.a.B. (1995). Genetics and Analysis of Quantitative Traits. Sunderland,MA: Sinauer

Associates, Inc. **980**.

Michaels, S.D., and Amasino, R.M. (1999). FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. The Plant Cell **11,** 949-956.

Miller, A.R., Scheereus, J.C., Erb, P.S., and Chandler, C.K. (1992). Enhanced strawberry seed germination through in vitro culture of cut achenes. Journal of the American Society for Horticultural Science **117,** 313-316.

Moon, J., Lee, H., Kim, M., and Lee, I. (2005). Analysis of flowering pathway integrators in Arabidopsis. Plant and Cell Physiology **46,** 292-299.

Morán, P., and Pérez-Figueroa, A. (2011). Methylation changes associated with early maturation stages in the Atlantic salmon. BMC genetics **12,** 1.

Morán, P., Marco-Rius, F., Megías, M., Covelo-Soto, L., and Pérez-Figueroa, A. (2013). Environmental induced methylation changes associated with seawater adaptation in brown trout. Aquaculture **392,** 77-83.

Morgan, H.D., Sutherland, H.G., Martin, D.I., and Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. Nature genetics **23,** 314-318.

Morris, J. (2001). Genes, genetics, and epigenetics: a correspondence. Science **293,** 1103-1105.

Mouhu, K., Kurokura, T., Koskela, E.A., Albert, V.A., Elomaa, P., and Hytönen, T. (2013). The Fragaria vesca homolog of SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1 represses flowering and promotes vegetative growth. The Plant Cell **25,** 3296-3310.

Mouhu, K., Hytönen, T., Folta, K., Rantanen, M., Paulin, L., Auvinen, P., and Elomaa, P. (2009). Identification of flowering genes in strawberry, a perennial SD plant. BMC plant biology **9,** 1.

Mouradov, A., Cremer, F., and Coupland, G. (2002). Control of flowering time interacting pathways as a basis for diversity. The Plant Cell **14,** S111-S130.

Mueller, U.G., and Wolfenbarger, L.L. (1999). AFLP genotyping and fingerprinting. Trends in Ecology and Evolution **14,** 389-394.

Murray, M.G., and Thompson, W.F. (1980). Rapid isolation of high molecular weight plant DNA. Nucleic Acids Research **8,** 4321-4326.

Nakao, M. (2001). Epigenetics: interaction of DNA methylation and chromatin. Gene **278,** 25-31.

Nitsch, J. (1970). Hormonal factors in growth and development. Biochemistry of fruits and their products, v. 1, 1970.

Oakeley, E.J. (1999). DNA methylation analysisa review of current methodologies. Pharmacology and Therapeutics **84,** 389-400.

Oakeley, E.J., and Jost, J.P. (1996). Non-symmetrical cytosine methylation in tobacco pollen DNA. Plant Molecular Biology **31,** 927-930.

Oosumi, T., Gruszewski, H.A., Blischak, L.A., Baxter, A.J., Wadl, P.A., Shuman, J.L., Veilleux, R.E., and Shulaev, V. (2006). High-efficiency transformation of the diploid strawberry (Fragaria vesca) for functional genomics. Planta **223,** 1219-1230.

Ossowski, S., Schneeberger, K., Lucas-Lledó, J.I., Warthmann, N., Clark, R.M., Shaw, R.G., Weigel, D., and Lynch, M. (2010). The rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana. science **327,** 92-94.

Parkin, I.A., Koh, C., Tang, H., Robinson, S.J., Kagale, S., Clarke, W.E., Town, C.D., Nixon, J., Krishnakumar, V., and Bidwell, S.L. (2014). Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. Genome biology **15,** 1.

Paszkowski, J., and Grossniklaus, U. (2011). Selected aspects of transgenerational epigenetic inheritance and resetting in plants. Current opinion in plant biology **14,** 195-203.

Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E., and Yanofsky, M.F. (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes. Nature **405,** 200-203.

Pelaz, S., Gustafson‐Brown, C., Kohalmi, S.E., Crosby, W.L., and Yanofsky, M.F. (2001). APETALA1 and SEPALLATA3 interact to promote flower development. The Plant Journal **26,** 385-394.

Peng, M., Cui, Y., Bi, Y.M., and Rothstein, S.J. (2006). AtMBD9: a protein with a methyl‐CpG‐binding domain regulates flowering time and shoot branching in Arabidopsis. The Plant Journal **46,** 282-296.

Pérez-Figueroa, A. (2013). msap: A tool for the statistical analysis of methylation-sensitive amplified polymorphism data. Molecular Ecology Resources **13,** 522-527.

Perkins‐Veazie, P. (2010). Growth and ripening of strawberry fruit. Horticultural Reviews, Volume 17**,** 267-297.

Poethig, R.S. (2003). Phase change and the regulation of developmental timing in plants. Science **301,** 334-336.

Portis, E., Acquadro, A., Comino, C., and Lanteri, S. (2004). Analysis of DNA methylation during germination of pepper (Capsicum annuum L.) seeds using methylation-sensitive amplification polymorphism (MSAP). Plant Science **166,** 169-178.

Potter, D., Luby, J.J., and Harrison, R.E. (2000). Phylogenetic relationships among species of Fragaria (Rosaceae) inferred from non-coding nuclear and chloroplast DNA sequences. Systematic Botany **25,** 337-348.

Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Tandem **60,** 0.8.

Rabinowicz, P.-T., ROCKVILLE, M., and Slovin, J. (2008). Frageria vesca, a useful tool for Rosaceae Genomics. In North American Strawberry Conference Proceedings, pp. 112.

Raj, S., Brautigam, K., Hamanishi, E.T., Wilkins, O., Thomas, B.R., Schroeder, W., Mansfield, S.D., Plant, A.L., and Campbell, M.M. (2011). Clone history shapes Populus drought responses. Proc Natl Acad Sci U S A **108,** 12521-12526.

Rakyan, V.K., Preis, J., MORGAN, H.D., and Whitelaw, E. (2001). The marks, mechanisms and memory of epigenetic states in mammals. Biochemical Journal **356,** 1-10.

Rakyan, V.K., Chong, S., Champ, M.E., Cuthbert, P.C., Morgan, H.D., Luu, K.V., and Whitelaw, E. (2003). Transgenerational inheritance of epigenetic states at the murine AxinFu allele occurs after maternal and paternal transmission. Proceedings of the National Academy of Sciences **100,** 2538-2543.

Ramchandani, S., Bhattacharya, S.K., Cervoni, N., and Szyf, M. (1999). DNA methylation is a reversible biological signal. Proceedings of the National Academy of Sciences of the United States of America **96,** 6107-6112.

Rapp, R.A., and Wendel, J.F. (2005). Epigenetics and plant evolution. New Phytologist **168,** 81-91.

Ratcliffe, O.J., and Riechmann, J.L. (2002). Arabidopsis transcription factors and the regulation of flowering time: a genomic perspective. Current issues in molecular biology **4,** 77-92.

Reinders, J., Wulff, B.B., Mirouze, M., Mari-Ordonez, A., Dapp, M., Rozhon, W., Bucher, E., Theiler, G., and Paszkowski, J. (2009). Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. Genes Dev **23,** 939-950.

Reyna-López, G.E., Simpson, J., and Ruiz-Herrera, J. (1997). Differences in DNA methylation patterns are detectable during the dimorphic transition of fungi by amplification of restriction polymorphisms. Molecular and General Genetics **253,** 703-710.

Richards, C.L., Bossdorf, O., and Pigliucci, M. (2010). What role does heritable epigenetic variation play in phenotypic evolution? BioScience **60,** 232-237.

Richards, E.J. (2011). Natural epigenetic variation in plant species: A view from the field. Current Opinion in Plant Biology **14,** 204-209.

Rico, L., Ogaya, R., Barbeta, A., and Penuelas, J. (2014). Changes in DNA methylation fingerprint of Quercus ilex trees in response to experimental field drought simulating projected climate change. Plant Biology **16,** 419-427.

Riggs, A.D., and Porter, T.N. (1996). Overview of epigenetic mechanisms. Cold Spring Harbor Monograph Archive **32,** 29-45.

Rival, A., Beulé, T., BERTOSSI, F.A., Tregear, J., and Jaligot, E. (2010). Plant epigenetics: from genomes to epigenomes. Notulae Botanicae Horti Agrobotanici Cluj-Napoca **38,** 09.

Roberts, R.J., Vincze, T., Posfai, J., and Macelis, D. (2005). REBASE - Restriction enzymes and DNA methyltransferases. Nucleic Acids Research **33,** D230-D232.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26,** 139-140.

Rousseau-Gueutin, M., Lerceteau-Köhler, E., Barrot, L., Sargent, D.J., Monfort, A., Simpson, D., Arús, P., Guérin, G., and Denoyes-Rothan, B. (2008). Comparative genetic mapping between octoploid and diploid Fragaria species reveals a high level of colinearity between their genomes and the essentially disomic behavior of the cultivated octoploid strawberry. Genetics **179,** 2045-2060.

Roux, F., Touzet, P., Cuguen, J., and Le Corre, V. (2006). How to be early flowering: an evolutionary perspective. Trends in plant science **11,** 375-381.

Roux, F., Colomé-Tatché, M., Edelist, C., Wardenaar, R., Guerche, P., Colot, V., Jansen, R.C., and Johannes, F. (2011). Genome-wide epigenetic perturbation jump-starts patterns of heritable variation found in nature. Genetics **188,** 1015-1017.

Russo, V.E., Martienssen, R.A., and Riggs, A.D. (1996). Epigenetic mechanisms of gene regulation. (Cold Spring Harbor Laboratory Press).

Salmon, A., Clotault, J., Jenczewski, E., Chable, V., and Manzanares-Dauleux, M.J. (2008). Brassica oleracea displays a high level of DNA methylation polymorphism. Plant Science **174,** 61-70.

Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F., and Coupland, G. (2000). Distinct roles of constans target genes in reproductive development of Arabidopsis. Science **288,** 1613-1616.

Sano, H. (2010). Inheritance of acquired traits in plants: reinstatement of Lamarck. Plant signaling & behavior **5,** 346-348.

Sano, H., Kamada, I., Youssefian, S., Katsumi, M., and Wabiko, H. (1990). A single treatment of rice seedlings with 5-azacytidine induces heritable dwarfism and undermethylation of genomic DNA. MGG Molecular & General Genetics **220,** 441-447.

Sargent, D., Fernandéz-Fernandéz, F., Ruiz-Roja, J., Sutherland, B., Passey, A., Whitehouse, A., and Simpson, D. (2009). A genetic linkage map of the cultivated strawberry (Fragaria× ananassa) and its comparison to the diploid Fragaria reference map. Molecular Breeding **24,** 293-303.

Saze, H. (2008). Epigenetic memory transmission through mitosis and meiosis in plants. Semin Cell Dev Biol **19,** 527-536.

Scheiner, S.M. (1993). Genetics and evolution of phenotypic plasticity. Annual review of ecology and systematics**,** 35-68.

Schmitz, R.J., and Ecker, J.R. (2012). Epigenetic and epigenomic variation in Arabidopsis thaliana. Trends in Plant Science **17,** 149-154.

Schmitz, R.J., Schultz, M.D., Lewsey, M.G., O'Malley, R.C., Urich, M.A., Libiger, O., Schork, N.J., and Ecker, J.R. (2011). Transgenerational epigenetic instability is a source of novel methylation variants. Science **334,** 369-373.

Schulz, B., Eckstein, R.L., and Durka, W. (2013). Scoring and analysis of methylation-sensitive amplification polymorphisms for epigenetic population studies. Molecular Ecology Resources **13,** 642-653.

Senanayake, Y., and Bringhurst, R. (1967). Origin of Fragaria polyploids. I. Cytological analysis. American Journal of Botany**,** 221-228.

Shan, X., Wang, X., Yang, G., Wu, Y., Su, S., Li, S., Liu, H., and Yuan, Y. (2013). Analysis of the DNA methylation of maize (Zea mays L.) in response to cold stress based on methylation-sensitive amplified polymorphisms. Journal of Plant Biology **56,** 32-38.

Shen, L., and Waterland, R.A. (2007). Methods of DNA methylation analysis. Current Opinion in Clinical Nutrition and Metabolic Care **10,** 576-581.

Shoemaker, J. (1978). Strawberries. Small Fruit Culture.

Shulaev, V., Korban, S.S., Sosinski, B., Abbott, A.G., Aldwinckle, H.S., Folta, K.M., Iezzoni, A., Main, D., Arus, P., Dandekar, A.M., Lewers, K., Brown, S.K., Davis, T.M., Gardiner, S.E., Potter, D., and Veilleux, R.E. (2008). Multiple models for Rosaceae genomics. Plant Physiol **147,** 985-1003.

Shulaev, V., Sargent, D.J., Crowhurst, R.N., Mockler, T.C., Folkerts, O., Delcher, A.L., Jaiswal, P., Mockaitis, K., Liston, A., Mane, S.P., Burns, P., Davis, T.M., Slovin, J.P., Bassil, N., Hellens, R.P., Evans, C., Harkins, T., Kodira, C., Desany, B., Crasta, O.R., Jensen, R.V., Allan, A.C., Michael, T.P., Setubal, J.C., Celton, J.M.,

Rees, D.J.G., Williams, K.P., Holt, S.H., Rojas, J.J.R., Chatterjee, M., Liu, B., Silva, H., Meisel, L., Adato, A., Filichkin, S.A., Troggio, M., Viola, R., Ashman, T.L., Wang, H., Dharmawardhana, P., Elser, J., Raja, R., Priest, H.D., Bryant, D.W., Fox, S.E., Givan, S.A., Wilhelm, L.J., Naithani, S., Christoffels, A., Salama, D.Y., Carter, J., Girona, E.L., Zdepski, A., Wang, W., Kerstetter, R.A., Schwab, W., Korban, S.S., Davik, J., Monfort, A., Denoyes-Rothan, B., Arus, P., Mittler, R., Flinn, B., Aharoni, A., Bennetzen, J.L., Salzberg, S.L., Dickerman, A.W., Velasco, R., Borodovsky, M., Veilleux, R.E., and Folta, K.M. (2011). The genome of woodland strawberry (Fragaria vesca). Nature Genetics **43,** 109-116.

Simon, R., Igeno, M.I., and Coupland, G. (1996). Activation of floral meristem identity genes in Arabidopsis. Nature **384,** 59-62.

Slovin, J.P., Schmitt, K., and Folta, K.M. (2009). An inbred line of the diploid strawberry Fragaria vesca f. semperflorens for genomic and molecular genetic studies in the Rosaceae. Plant Methods **5,** 15.

Smeets, L. (1980). Effect of temperature and daylength on flower initiation and runner formation in two everbearing strawberry cultivars. Scientia Horticulturae **12,** 19-26.

Song, Q.X., Xiang, L., Li, Q.T., Chen, H., Hu, X.Y., Ma, B., Zhang, W.K., Chen, S.Y., and Zhang, J.S. (2013). Genome-Wide analysis of DNA methylation in soybean. Molecular Plant **6,** 1961-1974.

Sønsteby, A., and Heide, O. (2007). Long-day control of flowering in everbearing strawberries. The Journal of Horticultural Science and Biotechnology **82,** 875-884.

Sønsteby, A., and Heide, O. (2008). Long-day rather than autonomous control of flowering in the diploid everbearing strawberry Fragaria vesca ssp. semperflorens. The Journal of Horticultural Science and Biotechnology **83,** 360-366.

Soppe, W.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M., and Peeters, A.J. (2000). The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. Molecular cell **6,** 791-802.

Souza, H.A., Muller, L.A., Brandão, R.L., and Lovato, M.B. (2012). Isolation of high quality and polysaccharide-free DNA from leaves of Dimorphandra mollis

(Leguminosae), a tree from the Brazilian Cerrado. Genetics and molecular research : GMR **11,** 756-764.

Springer, N.M. (2013). Epigenetics and crop improvement. Trends in Genetics **29,** 241-247.

Sridhar, V.V., Kapoor, A., Zhang, K., Zhu, J., Zhou, T., Hasegawa, P.M., Bressan, R.A., and Zhu, J.-K. (2007). Control of DNA methylation and heterochromatic silencing by histone H2B deubiquitination. Nature **447,** 735-738.

Stam, M. (2009). Paramutation: a heritable change in gene expression by allelic interactions in trans. Molecular plant **2,** 578-588.

Stauct, G. (1988). The species of Fragaria, their taxonomy and geographical distribution. In International Strawberry Symposium 265, pp. 23-34.

Steele, E.J., Lindley, R.A., Blanden, R.V., and Blanden, R. (1999). Lamarck's signature: how retrogenes are changing Darwin's natural selection paradigm. (Basic Books).

Stewart, P.J., and Folta, K.M. (2010). A review of photoperiodic flowering research in strawberry (Fragaria spp.). Critical reviews in plant science **29,** 1-13.

Stresemann, C. (2008). Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. International Journal of Cancer **123,** 8.

Sung, S., and Amasino, R.M. (2004). Vernalization in Arabidopsis thaliana is mediated by the PHD finger protein VIN3. Nature **427,** 159-164.

Susan, J.C., Harrison, J., Paul, C.L., and Frommer, M. (1994). High sensitivity mapping of methylated cytosines. Nucleic acids research **22,** 2990-2997.

Suter, L., and Widmer, A. (2013). Environmental heat and salt stress induce transgenerational phenotypic changes in Arabidopsis thaliana. PLoS One **8,** e60364.

Tanino, K.K., and Wang, R. (2008). Modeling chilling requirement and diurnal temperature differences on flowering and yield performance in strawberry crown production. HortScience **43,** 2060-2065.

Taylor, D., Atkey, P., Wickenden, M., and Crisp, C. (1997). A morphological study of flower initiation and development in strawberry (Fragaria x ananassa) using cryo‐scanning electron microscopy. Annals of applied biology **130,** 141-152.

Teixeira, F., and Colot, V. (2010). Repeat elements and the Arabidopsis DNA methylation landscape. Heredity **105,** 14-23.

Teper-Bamnolker, P., and Samach, A. (2005). The flowering integrator FT regulates SEPALLATA3 and FRUITFULL accumulation in Arabidopsis leaves. The Plant Cell **17,** 2661-2675.

Thu, K.L., Vucic, E.A., Kennett, J.Y., Heryet, C., Brown, C.J., Lam, W.L., and Wilson, I.M. (2009). Methylated DNA immunoprecipitation. JoVE (Journal of Visualized Experiments), e935-e935.

Torti, S., Fornara, F., Vincent, C., Andrés, F., Nordström, K., Göbel, U., Knoll, D., Schoof, H., and Coupland, G. (2012). Analysis of the Arabidopsis shoot meristem transcriptome during floral transition identifies distinct regulatory patterns and a leucine-rich repeat protein that promotes flowering. The Plant Cell **24,** 444-462.

Tran, R.K., Henikoff, J.G., Zilberman, D., Ditt, R.F., Jacobsen, S.E., and Henikoff, S. (2005). DNA methylation profiling identifies CG methylation clusters in Arabidopsis genes. Current Biology **15,** 154-159.

Van Kleunen, M., and Fischer, M. (2003). Effects of four generations of density‐dependent selection on life history traits and their plasticity in a clonally propagated plant. Journal of Evolutionary Biology **16,** 474-484.

Verhoeven, K.J., Jansen, J.J., van Dijk, P.J., and Biere, A. (2010). Stress-induced DNA methylation changes and their heritability in asexual dandelions. New Phytol **185,** 1108-1118.

Verhoeven, K.J.F. (2010). Changes in genomic methylation patterns during the formation of triploid asexual dandelion lineages. Molecular Ecology **19,** 315.

Verkest, A., Byzova, M., Martens, C., Willems, P., Verwulgen, T., Slabbinck, B., Rombaut, D., Van de Velde, J., Vandepoele, K., and Standaert, E. (2015). Selection for improved energy use efficiency and drought tolerance in canola results in distinct transcriptome and epigenome changes. Plant physiology **168,** 1338-1350.

Veselý, J. (1985). Mode of action and effects of 5-azacytidine and of its derivatives in eukaryotic cells. Pharmacology & therapeutics **28,** 227-235.

Veselý, J., and Čihák, A. (1978). 5-Azacytidine: mechanism of action and biological effects in mammalian cells. Pharmacology & Therapeutics. Part A: Chemotherapy, Toxicology and Metabolic Inhibitors **2,** 813-840.

Viegas-Péquignot, E., and Dutrillaux, B. (1976). Segmentation of human chromosomes induced by 5-ACR (5-azacytidine). Human genetics **34,** 247-254.

Visser, T. (1964). Juvenile phase and growth of apple and pear seedlings. Euphytica **13,** 119-129.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M. (1995). AFLP: A new technique for DNA fingerprinting. Nucleic Acids Research **23,** 4407-4414.

Waddington, C.H. (1942). Canalization of development and the inheritance of acquired characters. Nature **150,** 563-565.

Wade, P.A. (2001). Methyl CpG‐binding proteins and transcriptional repression*. Bioessays **23,** 1131-1137.

Wang, R., Farrona, S., Vincent, C., Joecker, A., Schoof, H., Turck, F., Alonso-Blanco, C., Coupland, G., and Albani, M.C. (2009). PEP1 regulates perennial flowering in Arabis alpina. Nature **459,** 423-427.

Wassenegger, M., Heimes, S., Riedel, L., and Sanger, H.L. (1994). RNA-directed de novo methylation of genomic sequences in plants. Cell **76,** 567-576.

Weber, M. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nature Genet. **37,** 853-862.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schuebeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nature genetics **37,** 853-862.

Wilson, D., Goodall, A., and Reeves, J. (1973). An improved technique for the germination of strawberry seeds. Euphytica **22,** 362-366.

Wilson, R.N., Heckman, J.W., and Somerville, C.R. (1992). Gibberellin is required for flowering in Arabidopsis thaliana under short days. Plant Physiology **100,** 403-408.

Wolffe, A.P., and Matzke, M.A. (1999). Epigenetics: Regulation through repression. Science **286,** 481-486.

Xi, Y., and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPping program. BMC bioinformatics **10,** 1.

Xu, M., Li, X., and Korban, S.S. (2004). DNA-methylation alterations and exchanges during in vitro cellular differentiation in rose (Rosa hybrida L.). Theor Appl Genet **109,** 899-910.

Yaish, M.W., Peng, M., and Rothstein, S.J. (2009). AtMBD9 modulates Arabidopsis development through the dual epigenetic pathways of DNA methylation and histone acetylation. The Plant Journal **59,** 123-135.

Yoo, S.K., Chung, K.S., Kim, J., Lee, J.H., Hong, S.M., Yoo, S.J., Yoo, S.Y., Lee, J.S., and Ahn, J.H. (2005). CONSTANS activates SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 through FLOWERING LOCUS T to promote flowering in Arabidopsis. Plant physiology **139,** 770-778.

Yu, Q., and Ma, H. (2001). The flowering transition and florigen. Current biology : CB **11**.

Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science **328,** 916-919.

Zhang, C., and Hsieh, T.-F. (2013). Heritable epigenetic variation and its potential applications for crop improvement. Plant Breeding and Biotechnology **1,** 307-319.

Zhang, K., Sridhar, V.V., Zhu, J., Kapoor, A., and Zhu, J.-K. (2007). Distinctive core histone post-translational modification patterns in Arabidopsis thaliana. PloS one **2,** e1210.

Zhang, L., Wang, Y., Zhang, X., Zhang, M., Han, D., Qiu, C., and Han, Z. (2012). Dynamics of phytohormone and DNA methylation patterns changes during dormancy induction in strawberry (Fragaria× ananassa Duch.). Plant cell reports **31,** 155-165.

Zhang, M., Xu, C., Yan, H., Zhao, N., vonwettstein, D., and Liu, B. (2009). Limited tissue culture-induced mutations and linked epigenetic modifications in F1 hybrids of sorghum pure lines are accompanied by increased transcription of DNA methyltransferases and 5-methylcytosine glycosylases. The Plant Journal: **57,** 666.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W.-L., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., and Jacobsen, S.E. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. Cell **126,** 1189-1201.

Zhang, Y.Y., Fischer, M., Colot, V., and Bossdorf, O. (2013). Epigenetic variation creates potential for evolution of plant phenotypic plasticity. New Phytologist **197,** 314-322.

Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., and Henikoff, S. (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nat Genet **39,** 61-69.